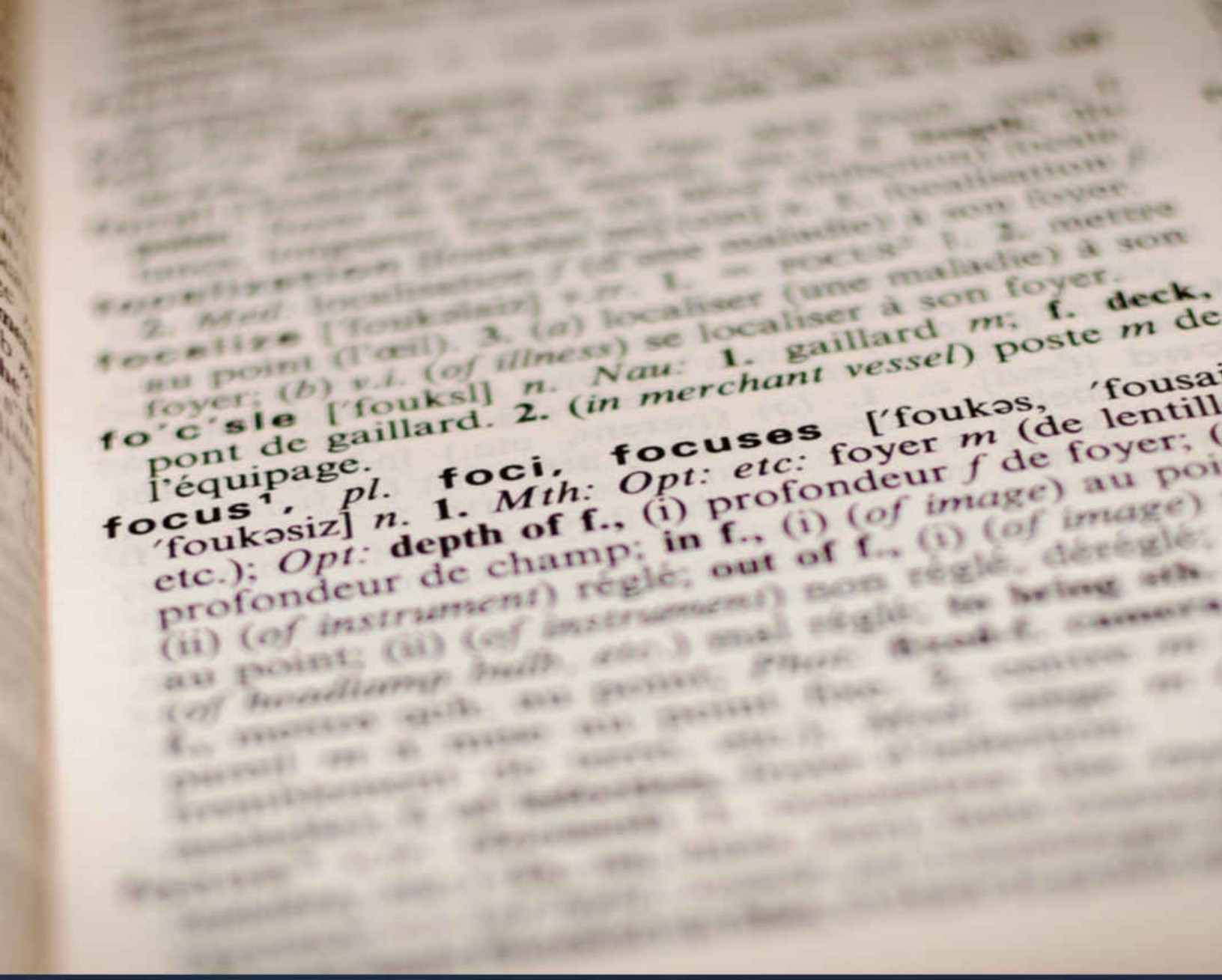

LITERAL MEANING & COGNITIVE CONTENT

BY J.–M. KUCZYNSKI

Literal Meaning & Cognitive Content

J-M Kuczynski, PhD

Summary of the Present work (August 12, 2015)
Preface (August 11, 2015)
Volume I: Literal Meaning
Chapter 1: Introduction to Volume I
Chapter 2 Meaning versus Reference
Chapter 3 The Concept of Logical Equivalence: Predicate-equivalence vs. sentential-equivalence vs. sentence-token equivalence vs. propositional equivalence
Chapter 4 Do Quantifiers refer to functions?
Chapter 5 Why sentences refer to propositions: a preliminary argument
Chapter 6 What do sentence-tokens refer to?
Chapter 7 Black's Principle and Church-Black Debate
Chapter 8 Are all terms referring terms?
Chapter 9 The confusions underlying the sense-reference distinction
Chapter 10 The causal theory of reference
Chapter 11 Are all terms referring terms?
Chapter 12 The semantics of non-indicatives
Chapter 13 Frege on indirect discourse and compositionality
Chapter 14 Opacity revisited: the inadequacy of Kaplan's analysis, the viability of other analyses
Chapter 15 Our semantics
Chapter 16 Truth as Instantiatedness and the Unity of the Proposition
Volume II : Cognitive Content
Chapter 17 Introduction to Volume II
Chapter 18 Extensionality and cognitive content
Chapter 19 Externalism and self-knowledge
Chapter 20 Literal versus communicated meaning
Chapter 21 Uniquely individuating descriptions and the problem of incomplete knowledge
Chapter 22 Program-causes and the Jackson-Pettit Attempt to Save Externalism
Chapter 23 Externalism and the Veil of Perception

Chapter 24 Empty names
Chapter 25 Fictitious names
Chapter 26 Anaphoric reference and the referential-attributive distinction
Chapter 27 Definite descriptions revisited
Chapter 28 Can reference be reduced to quantification?
Bibliography

Preface (August 11, 2015)

The present work concerns the relationship between word-meaning and thought-content. Volume I of this two volume work concerns word-meaning, and Volume II concerns thought-content.

Summary of Volume I: It is made clear what it is for a given linguistic expression E to have a given meaning M, and it is thereby made clear exactly how a given sentence's meaning depends on this meanings of its parts is made. Further, the Fregean thesis that 'grammatical surface-structure may diverge from logical form' is evaluated.

It is shown that Frege's thesis is correct only in relation to erroneous views as to the semantics of quantifiers and intensional operators. It is clearly stated what such expressions actually mean, and on this basis it is shown that, *pace* Frege, logical and grammatical form do in fact coalesce---as they obviously must, given that, after allowances are made for the relatively small number of expression-specific linguistic rules that one can learn, sentences are intelligible only to the extent that surface-structure mirrors logical form.

Most importantly, it is said exactly what linguistic meanings are; and on this basis it is clearly stated how a complex (composite) expression's grammatical structure maps onto the structure of the meaning encoded in that expression.

On this basis, a demonstrably correct solution is put forth to Frege's paradox concerning the "unity of the proposition." This paradox is given by the *prima facie* cogent point that, because no string of nouns constitutes a sentence (except in relation to linguistic conventions that assign phonetically unrealized semantic structure to what would otherwise be strings of nouns) , it follows that, for any admissible values of "*a"* and "*PHI"*, the occurrence in "*a has PHI"* of "*PHI*" does not refer to *PHI* . (Explanation "Smith runs fast" is meaningful, but "Smith, the property of running fast" is not meaningful, the

latter being meaningless expression whole that contains meaningful parts.

Summary of Volume II: It is made clear what it is for a given person (or creature of any kind) have a concept of a given object. Russell-Frege descriptivism is proven correct; i.e. it is shown that person for person S to have concept of object O is for S to know some truth of the form *for some x, x has* $PHI_1 ... PHI_n$, where O and O alone has all of $PHI_1 ... PHI_n$. In others words, for S to have a concept of O is for S to know some truth that singles O out. But this is only a necessary, as opposed to a sufficient, condition for S's having knowledge of O: it is also necessary that $PHI_1 ... PHI_n$ be such that they relativize O to S's cognitive map, meaning that, in virtue of knowing the truth *for some x, x has* $PHI_1 ... PHI_n$, S has knowledge of a truth that not only singles out O, but singles out O in a way in which it is cannot possibly be singled out for any other creature. (This important truth is neither stated nor defended in the present work, but it is stated and thoroughly defended in *Conceptual Atomism and the Computational Theory of Mind*, written one year after the present treatise.)

The tiny kernel of truth in the otherwise disastrously misconceived 'photograph-theory' of conception (viz. the relation between S and O, where S has a concept of O, is the same as the relation between a photograph and the thing photographed) is that, when S has a concept of O, it is partly because S Burge-style externalism is shown to embody a failure to distinguish between word-meaning and thought content. And it is shown that given a correct understanding of the distinction between the two, the relevant data can modeled in such a way as to validate our pretheoretic intuitions as well as the presumptions of empirical science.

The model put forth, and vociferously defended, by Tyler Burge does not satisfy these desiderata. Specifically, Burge's analysis entails that:

> (i) Thoughts do nothing: the mental sphere is causally inert (and therefore non-existent, given that, where spatiotemporal entities are concerned, to exist is to have causal properties);
>
> (ii) If S has a belief about O, where O is some object in the external world (as opposed to an abstract entity, e.g. a number, or a sensation, e.g. a tickle, experienced by S), S cannot know himself to have that belief;

(iii) The mind is, quite literally, spread out over space-time: if you have a concept of a star that saw through a telescope but that went out of existence long-ago, it is because that star is an actual constituent---not an object, be it emphasized, but a component---of your mind.

That (iii) is indeed a consequence of Burge's position was proved in by Colin McGinn (1988).

Of course, these consequences of Burge's theory are totally absurd. Content-externalists have spent the last 36 years trying to show that (i)-(iii), along with the various other absurdities that follow from Burge's theory, are either true (which they are not) or that they do not follow from Burge's theory (which they do).

Burge's theory, as well as its various derivatives---e.g. Fodor's atomism about concepts (viz. one's concept of the number 7 has no logical or psychological complexity), which is an immediate consequence of the photograph-theory, which in turn is an immediate consequence, if not the very essence, of Burge's theory---are in each and every case shown to embody a failure to distinguish between literal meaning and cognitive content. And this failure is in its turn rooted in two failures:

(i) A failure to understand literal meaning, and
(ii) A failure to understand cognitive content.

This book explains

(1) What literal meaning is,
(2) What cognitive content is,
(3) How they differ, and
(4) How they are related.

General Points: This book is longer than it should be. (I wrote it when I was in graduate school, when I had a tendency to over-explain and also to over-focus on technicalities.)

That said, most of the chapters in the present work are self-contained (even though they do collectively constitute a single, organic entity); and if a

reader wishes to read the material on, for example, the structure of the proposition (Chapter 16) or the nature of discourse about fictitious entities (Chapter 25), he can skip right to the corresponding chapter.

Apart from the defense of Montague-grammar (Chapter 15) and the analysis of the proposition (Chapter 16), the material in Volume I is likely to be of little interest except to hyper-competent scholars who specialize in this particular subdiscipline. It addresses problems that, at the time I wrote the present treatise, were of great concern to the philosophical community (not in the sense that they were concerned with, or able to ascertain, the truth about those problems, but in the sense that they were the foils-du-jour, whereby a quickly degenerating philosophical community attempted to appear to be philosophizing). And the reason they were of such concern is, I am afraid, that the average intelligence-level of philosophers in 2005, though no lower than it is now, was inappropriately low. As a result, philosophical discourse, especially in semantics and philosophical logic, consisted of bitter debates concerning technical problems that were internal to the broken theories of the participants of those debates and were misunderstood even within those parameters.

In Part I, Chapters 1-10, those debates are adjudicated. After thus clearing the way of this philosophical debris, I proceed to deal with problems of real substance, for example:

(i) The nature of meaning and truth (Chapter 16)
(ii) The nature of conception (Chapter 18), and
(iii) The semantics of fictitious statements (Chapters 25-26).

The contents of Chapters 1-10 are concerned with the very same problems as

*Keith Donnellan’s article "Reference and Quantification", the central thesis of which was plagiarized *verbatim* from A.J. Ayer’s (1956) book *The Problem of Knowledge*,

*Kripke’s (1969) *Naming and Necessity*, which is little more than an exposition (*sans* acknowledgement of their progenitors) of points made in the late 50s and early 60s by Arthur

Pap (whose works, suggestively, were blacklisted in the late 60s, after he had died) and, to a lesser extent, some (early 60s) papers on modal logic by Ruth Marcus,

*David Kaplan's (1989) *Demonstratives*, which is nothing more than an exposition, *sans* acknowledgement of their progenitor, of ideas put forth in the 50s by Peter Strawson (who, being a genuine philosopher, had broad interests and who, suggestively, distanced himself from anything having to do with the philosophy of language at about the time (late 60s) that Kripke and Kaplan were becoming prominent),

*Michael Dummett's (1973) book *Frege: The Philosophy of Language*, which is one of the great philosophical works of the 20th century, and

*Joseph Almog's (2014) *Referential Mechanics*.

In Part II, the semantic arcana dealt with in Part are integrated into an original epistemological system, the legitimacy of which is then established on independent grounds.

Volume I: Literal Meaning
Chapter 1: Introduction to Volume I

"Snow" refers to snow. "Nixon" refers to Nixon. In this work, it is shown that given a correct understanding of the nature of reference and of linguistic meaning generally, it is possible to produce non-revisionist analyses of the nature of

*Perceptual content,
*Mental content generally,
*Logical equivalence,
*Logical dependence generally,
*Counterfactual truth,
*The causal efficacy of mental states,

and

*Our knowledge of ourselves and of external world.

In addition, set-theoretic interpretations of several semantic concepts are put forth. These concepts include *truth, falsehood, negation,* and *conjunction.*

Given the doctrines put forth in this work, many puzzles of a logico-semantic nature are readily solved. One class of puzzles concerns statements about non-entities (e.g. the positive square root of negative two). Another such class concerns identity-statements (e.g. 2 is the predecessor of 3). In this work, it is said what such statements mean and why they mean it.

Our analyses of these statement-classes are compatible with Kripke's cogent (1969) defense of Direct Reference Theory (*names pick out their referents by labelling them, not by describing them*), and they show that theory to be consistent with classical logic.

Many semanticists deny this last assertion and on that basis advocate far-reaching revisions of the laws of logic and commonsense. For example, according to Nathan Salmon and Scott Soames, Direct Reference Theory that:

(i) Socrates simultaneously exists and fail to exist,

and also more generally that:

(ii) A given state of affairs can simultaneously obtain and fail to obtain.

Being direct reference theorists themselves, Salmon and Soames accept (i) and (ii), and they also hold that:

(iii) The laws of logic must be revised, so as to accommodate the putative truth of (i) and (ii).

Salmon also believes, since he believes it to be a consequence of Direct Reference Theory, that:

(iv) To be is to conceived (e.g. given that one can have thoughts about perpetual motion machines, such machines exist).

and also that:

(v) One can rationally accept statements of the form *P and not-P* (e.g. Jim has a car and Jim does not have a car)

Salmon duly holds that:

(vi) The laws of classical logic must be revised, so as to accommodate the putative truth of (iv) and (v).

In the present work it is shown that Direct Reference Theory has none of these consequences, any appearance to the contrary being rooted in a failure to understand the relationship between the literal meanings of our words and contents of the thoughts embodied therein.

This same failure, it is shown, is the very essence of Burge's contention that:

(vii) Should it turn out that Moses never existed, your Moses-thoughts would turn out to be non-thoughts,

and more generally that:

(viii) The content of a given thought is irrelevant to its causal properties.

A corollary of (viii) is that:

(ix) A given thought's psychological properties are unrelated to its content.

A corollary of (ix) is that

(x) Thoughts do nothing.

Given that no occupant of space-time is without causal properties, a consequence of (x) is that:

(xi) Thoughts do not exist.

Thus, (viii) entails (xi). Since (xi) is false, so is (viii).

Burge denies that (viii) entails (xi). He defends this position by taking the position that:

(xii) An object's causal properties cannot be understood in terms of events in its immediate neighborhood.

This move on Burge's part is contraindicated in many ways.

First of all, (xii) is *ad hoc*.

Further, (xii) is incompatible with basic precept of modern mechanics: *No action at a distance*, which, even if not strictly true, is accurate to a higher degree of approximation than is consistent with Burge's theory.

Moreover, considerations of a purely philosophical kind cannot possibly warrant rejection of that precept. Nor *a fortiori* can such considerations warrant rejection of the corresponding generalization: *Nothing happens if could just as well not happen*, known as the *Principle of Sufficient Reason.*

Finally, given that philosophical thought is only as cogent as the Principle of Sufficient Reason is true, Burge's theory is not only false but self-undermining.

Being a textbook case of a degenerating research program, Burge's theory and its derivatives, which include the previously mentioned theories of Nathan Salmon's, have duly given way to an epicycle-free analysis, which it is the purpose of the present volume to state and defend.

The semantic doctrines defended by the present work

Let us now state some of the semantic (meaning-related) doctrines that it is the purpose of the present work to refute:

(i) Some expressions have both meaning and reference. "The inventor of bifocals" has the same reference as "the first post-master general". But those expressions differ in meaning (or "sense").

(ii) An expression can be meaningful without referring to anything. Words like "or", "not", and "and" do not refer to anything. But they are obviously significant. The same is true of morphemes like "-s" and "-ed".

(iii) The meaning of a sentence is a *proposition.* Consider the sentence "snow is white". The *meaning* of this sentence is identical with the *referent* of the expression "the proposition that snow is white".

Here is part of what I will try to establish in this work:

(i*) No entity has both meaning and reference. Reference is a property of spatio-temporal entities (expression tokens). Sense is a property of platonic entities (expression types).

(ii*) Tokens of "or", "if", and the like *do* refer, no less than tokens of "Plato" and "Socrates". Given a correct conception of reference, the difference between "categorematic" and "syncategorematic" entities collapses. Expressions that *seem* not to refer to anything – e.g. case-markers, sentence-connectives – refer to higher-order functions. The meaning of an expression-token always coincides with its referent. No meaningful expression-token does *not* refer.

(iii*) Sentence-types do not have propositions for the meanings, and neither do sentence-tokens. The meaning of a sentence-token is a certain kind of function *involving*, but not identical with, a proposition. The meaning of a sentence-type is a function involving, but not identical with, the kind of function just mentioned.

There are two other views that I wish to combat. These two conflict with each other, but each has its proponents:

(iv) Sentences refer to truth-values.

(v) Sentences don't refer to anything.

First of all, there are sentence-types and sentence-tokens. There is no such a thing as a sentence *simpliciter.* So (v) and (vi) are ambiguous. (v) is false on both disambiguations. (vi) is false on one, true on another. What I believe is:

(iv*) Sentence-*types* do not refer to anything: indeed, no expression-*type* refers. Reference is a property of expression-tokens. Sentence-*tokens* do not refer to truth-values.

(v*) Sentence-tokens *do* refer. They refer to their meanings. (Very roughly, the meaning of a sentence-token is a proposition; less roughly, it is a certain kind of function involving a proposition.)

The misunderstandings that dominate contemporary philosophy of language have crystallized in an argument known as the "Slingshot". The Slingshot "proves" things that are obviously not true (e.g. that your car is identical with my computer). So we must look for errors in it. These are not hard to find. Once these errors are exposed, it becomes necessary to rethink and reconstruct many semantic orthodoxies. The purpose of this work is to effect that reconstruction.

If cogent, the narrowly semantic arguments given in this work have far-reaching implications. They give us leverage on some perennial issues concerning knowledge and psychological explanation.

The structure of the present work

The present work is divided into two major sections or "volumes", as I will call them. In the first volume (Chapters 2 through 16), I defend various semantic claims, the main ones being (i*)-(iv*). That volume will be concerned almost exclusively with semantics. Virtually nothing will be said about epistemology, modality, or psychological explanation.

In the second volume (Chapters 17 through 28), I draw what I believe to be the non-semantic consequences of the narrowly semantic points made in the first part. I discuss how, in light of those narrowly semantic points, we can obtain some insight into various problems relating to a doctrine (or, better, a family of interconnected doctrines) called "externalism". In connection with this, we will obtain insight into problems relating to epistemology, modality, and psychological explanation.

Up to a point, the first volume is an articulation and consolidation of existing views. It will inevitably arouse disagreement. But because of its conservative nature, I don't think that volume will be *particularly* controversial.

The second volume is a different matter. There I *will* try to overthrow many orthodoxies of the present day. So the second volume, unlike the first, will arouse fierce and widespread resistance, though I will do my best to mitigate that through argumentation.

The second volume will not be sheer iconoclasticism. I will indeed try to overthrow some contemporary orthodoxies. But I will do so by appealing to older and more deeply entrenched ones. So, from some viewpoint, what I say is actually quite conservative.

I should point out that, in Volume II, I revisit certain narrowly semantic issues (e.g. whether anaphoric pronouns are bound variables) in order to strengthen the non-semantic points for which I am there arguing. So the two volumes are intertwined in subject-matter.

In what remains of this chapter, I will give a more precise statement of the views that will be defended in this work; and I will also give a brief *outline* my arguments for those views. But the arguments proper will be left to subsequent chapters.

Chapter 2 Meaning versus Reference

At some level, every version of the Slingshot assumes:

(SS) Replacing a referring term with a co-referring term can result in a change of meaning. Right now I want to give a couple of illustrations of how the Slingshot depends on that principle. And then I want to show why, in my judgment, that principle is false.

Church 1956

Let us assume:

(CR) Replacing expressions with co-referring expressions cannot result in a change of *reference.*

(SR) Expressions that are synonymous (or "very nearly so") co-refer.

According to Church (1956), if we grant these principles, then we can prove that sentences refer to truth-values:

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is the author of Waverly.
(iii) Sir Walter Scott is the man who wrote twenty-nine Waverly novels altogether.
(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) The number of counties in Utah is twenty-nine.

By CR, (i) and (ii) and (iii) must all refer to the same thing.
(iii) and (iv) are, in Church's view, synonymous (or "very nearly so").
Nearly synonymous expressions must co-refer. So (iii) and (iv) co-refer.
By CR, (v) refers to the same thing as (iv).
It follows that (i) and (v) refer to the same thing. They also have the same truth-value. Indeed, the one significant thing that (i) and (v) have in common is that they are both true.
CR preserves reference. The one significant property of sentences

that CR preserves is truth-value. So sentence-reference is sentence truth-value.

What is true of (i) and (v) is true (mutatis mutandis) of *any* two true sentences. Let S and S* be any two true sentences you choose. Given CR, plus the innocuous principle that synonymous expressions co-refer, a little bit of syntax-chopping shows that S and S* have the same referent.

Thus any two true sentences refer to the same thing: the property of truth.

An exactly similar argument shows that any two false sentences refer to the same thing: the property of falsity.

A problem

There is a problem with Church's argument. Church assumes that "Scott" and "the author of Waverly" both refer to Scott. Suppose he is right. In that case,

(i) Scott was wise

and

(ii) The author of Waverly was wise

ought to mean the same thing. They both ought to mean:

(iii) *Scott was wise.*

What does it mean to say that "the author of Waverly" refers to Scott? Presumably, it means exactly this: sentences of the form "…the author of Waverly…" mean:…*Scott.*…

Therefore, replacing "the author of Waverly" ought not to result in a change of meaning.

So lines (i) and (ii) must *both* mean:

(i_R) *Scott is Scott.*

Church assumes that "twenty-nine" refers to the same thing as "The number of counties in Utah" and also "the number n, such that Sir Walter Scott wrote n Waverly novels". He assumes they all refer to the number twenty-nine.

Let us suppose he is right. In that case, by reasoning exactly similar to that just given, lines (iii)-(v) must all mean:

(iii_R) The number twenty-nine is identical with the number twenty-nine.

Thus (i)-(v) are synonymous with:

(i_R) *Scott is Scott.*
(ii_R) *Scott is Scott.*
(iii_R) *Scott is Scott.*
(iv_R) The number twenty-nine is identical with the number twenty-nine.
(v_R) The number twenty-nine is identical with the number twenty-nine.

(iii_R) and (iv_R) are not synonymous. The argument fails.

The response: SS

There is an obvious response. Church can say:

Two terms can co-refer and have *different meanings.* "Scott" and "the author of Waverly" both refer to Scott. But

(i) "Scott was wise"

and

(ii) "The author of Waverly was wise"

obviously have different meanings. We explain this by saying: referring terms have both sense and reference. What a referring term semantically contributes is not *just* its referent. And co-referring terms can have different senses.

If this is right, then terms that co-refer can make *different* semantic contributions. This is equivalent to saying: in some cases, replacing a referring term with a co-referring term *can* result in a change of "sense" or meaning. Let us refer to this principle as (SS).

(SS) Replacing a referring term with a co-referring term can result in a change of meaning.

If (SS) is false, then Church's argument is blocked. So Church's argument does not assume only (LL) and (CR). It also assumes the correctness of a certain semantic theory: it assumes that, at least in some cases, co-referring terms can have different meanings and, consequently, that replacing a referring term with a co-referring term can result in a change of meaning.

Frege's argument

Frege's original argument[1] assumes (CR).
Consider the following two sentences:

(a) "Socrates was wise".
(b) "The unique x such that triangles have three sides and x drank hemlock and x=Socrates was wise".

(b) is what results when a referring term in (a) is replaced with a co-referring term.
By (CR), (a) and (b) cannot differ in their *referent.*
The fact that (a) and (b) have the same referent seems to show that the referent of a sentence is *not* the proposition it encodes. For it seems quite plain that (a) encodes a different proposition from (b). (b) encodes a

proposition about hemlock-drinking; (a) does not.

So given that (a) and (b) must have the same referent, and given *also* that they encode different propositions, it *does* follow that what a sentence refers to is not the proposition it bears.

Given only that sentences do not refer to propositions, it does not yet follow that sentences refer to truth-values. But, on the basis of what has been said, Frege is able to show just this.

Let S be any true sentence, and let P be any true proposition. Merely by replacing referring terms with co-referring terms, we can produce a sentence S* that has P for at least part of its meaning.

Pick any true proposition you wish – say, *arithmetic is incomplete*. Pick any sentence you wish – say (a). Merely by replacing a referring term in (a) with a co-referring term, we can produce a sentence that has the proposition *arithmetic is incomplete* for at least part of its meaning:

(c) "the unique x such that arithmetic is incomplete and x=Socrates was wise."

This point generalizes without limit. Take any true sentence you wish. Call it S. Take any true proposition you wish. Call it P. If we allow ourselves *only* to replace the referring terms in S with co-referring terms, we can produce a true sentence S* at least part of whose meaning is P.

By CR, S and S* must have the same *referent.*

S and S* needn't mean the same thing. In fact, there is no limit to how much they may differ in respect of what their meanings are.

Indeed, if we allow ourselves only to replace referring terms with co-referring terms, we can destroy *almost* everything that is semantically significant about a sentence.

But the *truth-value* will never be changed.

CR does *not* preserve sense.

CR *does* preserve reference.

CR preserves truth-value.

Reference is the *only* thing CR preserves.

Therefore, sentence-reference must be sentence truth-value. If sentences refer to anything, then all true sentences must refer to the truth-value *true*.

An exactly analogous argument shows that all false sentences must

refer to the truth-value *false.*

Let us sum up. Given CR, it seems to follow that all true sentences refer to the same thing – the truth-value *true* – and all true sentences refer to the same thing – the truth-value *false.*

Frege's need for (SS)

Frege's argument, like Church's, implicitly assumes (SS).

If (a) and (b) mean the *same* thing, we have no reason to believe that the referent of a sentence is anything other than the proposition it encodes.

(b) is what results when a referring term in (a) is replaced with a co-referring term. If that replacement does *not* result in a change of meaning, then there is no reason to think that the referent of (a) or (b) is anything other than its meaning.

In general, if replacements of referring terms with co-referring terms do not result in changes of meaning, then there is no reason to believe that co-referring sentences can differ in meaning. Replacing referring terms with co-referring terms surely preserves sentence-reference. But if it also preserves meaning, then sentence-meaning and sentence-reference never pull apart, and we have *no* reason for believing that sentences refer to anything other than their meanings.

But if we accept Frege's assumptions, then it seems that (a) and (b) *must* encode the same proposition. Frege assumes that "Socrates" refers to the same thing as "the unique x such that x drank hemlock and x=Socrates". In that case, those expressions ought to contribute the exact same thing (Socrates). So (b) ought to be synonymous with:

(b_R) "Socrates was wise."

By exactly similar reasoning, (c) ought to be synonymous with:

(c_R) "Socrates was wise."

But (c_R) and (b_R) encode precisely the same proposition as (a). Contrary to what Frege says, it turns out *not* to be possible to change sentence-meaning by replacing referring terms with co-referring terms. This

destroys his argument.

Frege has a famous response. He says that referring terms have both "sense" and reference.[2] A referring term contributes not only its referent but its sense. Co-referring terms can have different senses. So it *is* possible to change sentence-meaning by replacing referring terms with co-referring terms.

Let us sum up. If it is *not* possible to change sentence-meaning by replacing referring terms with co-referring terms, then Frege's argument is blocked: all of (a)-(c) encode the very same proposition. If it isn't possible to change what a sentence means by replacing referring terms with co-referring terms, then there is no longer any reason to suppose that a sentence's referent is different from its meaning. So Frege needs (SS) to be true.

More on SS

I will argue that SS is simply a confusion: actually a mass of confusions. It involves confusions of expression-types with expression-tokens. It involves confusions of literal meaning and communicated meaning – more exactly of semantics with *pre*-semantics.

But SS *seems* extremely plausible. In fact, its negation seems extremely implausible. Surely "Plato" and "the greatest teacher of Aristotle" co-refer. Yet surely they have different meanings. So surely

"Plato was wise"

doesn't have quite the same meaning as

"the greatest teacher of Aristotle was wise".

No one can deny the force of this point.

Some of the incoherencies in Frege's view

Let us begin by stating the more important components of Frege's semantics.

"The inventor of bifocals" is a singular term. But when that expression occurs in a sentence, the proposition thereby meant does *not* have

Benjamin Franklin as a constituent. Rather, it has the sense of the expression "*inventor of bifocals"* as a constituent.

Frege also holds that when an expression occurs in an "oblique" context, it undergoes a shift in reference. Right now, let us illustrate the concept of obliquity with some examples.

(*) "John believes that the inventor of bifocals snored".

(*) comprises the sentence "the inventor of bifocals snored". But the truth-value of the (*) is not a function of the truth-value the component sentence. If you replace "the inventor of bifocals snored" with another true sentences (e.g. "there are infinitely many primes") the resulting host sentence may be false.

The occurrence of "the inventor of bifocals snored" constitutes a piece of what Latinists refer to as "indirect discourse". It is not occurring assertively in that context. Contrast (*) with:

(**) "The inventor of bifocals snored and snow is white".

In (**), both component sentences do occur assertively.

"Oblique context" is usually defined thus. Let s be a sentence that occurs as a part of some other sentence S. s occurs non-obliquely if it occurs truth-functionally, and it occurs obliquely if it occurs non-truth-functionally.

But this definition isn’t accurate:

(***) "It is true that the inventor of bifocals snored."

In (***), "the inventor of bifocals snored" occurs truth-functionally but obliquely. Why do I say this? In (***), "that the inventor of bifocals snored" is doing exactly the same thing that it is doing in (*). Therefore the same is true of "the inventor of bifocals snored". So an oblique occurrence may be truth-functional, contrary to what is usually maintained.

Right now, we are going to set aside the question how exactly to define "oblique context." We will soon revisit it, to discover that on any legitimate definition of "oblique context", Frege’s system collapses. But right now we need only register one of Frege’s more startling positions. He holds

that, when a referring term occurs in an oblique context, it doesn't refer to what it otherwise refers to: rather, it refers to its sense. So in both (*) and (***), "the inventor of bifocals" doesn't refer to Franklin; it refers to the concept *inventor of bifocals.* And in

(****) "It is true John believes that the inventor of bifocals snored",

it refers to a *concept* of that concept.

Frege had a very good reason for holding this view. There is every reason to accept compositionality. There is also every reason to believe that

(bfs) "That the inventor of bifocals snored"
does *not* refer to the same thing as

(pmgs) "That the first post-master general snored".

They seem to refer to different propositions.

Finally, there is every reason to believe that "the inventor of bifocals" and "the first post-master general" are co-referring terms.

With one qualification, Frege held all these views.

Frege also realized that they cannot all be consistently held. What (a) refers to is a function of the *sense* of the definite description occurring in it. The same is true of (b). So compositionality requires Frege to hold that in those contexts, definite descriptions refer to their customary senses. For similar reasons, he holds that the same thing applies to proper names. So in oblique contexts referring terms refer to their senses, not their ordinary referents.

§ Frege had another powerful reason for supposing that, in oblique contexts, referring terms refer to their ordinary sense. These reasons relate to Leibniz's Law. If A=B, then A has a property P iff B has that property.

Throughout this work, when I talk about Leibniz's law, I am not talking about a principle governing the intersubstitutibility of *expressions.* I am talking about a rule concerning *objects.* If the *object* A really is identical with the object B, then surely A cannot have any property not had by B. If A really is such that John believes it to be equal to two, then the same is true of B.

Frege believed that, in some contexts, replacing a referring term with a co-referring term changes truth-value:

(A) "John believes that two is greater than one." (True)

(B) "John believes that the unique number n such that n=2 if there are continuous functions that cannot be differentiated at any point and such that n=1 otherwise is greater than one." (False.)

It is not possible that the number two should both have and lack a given property. Either it has it or it doesn't. If the number one has the property of being such that John believes it to be greater than one, then it has that property. Period. If the number one has that property, then anything identical with it has that property. So that property is also had by the unique number n such that n=2 if there are continuous functions that cannot be differentiated at any point and such that n=1 otherwise is greater than one.

But there is no doubt that the proposition communicated by (A) has a different truth-value from that communicated by (B). (And it seems highly reasonable – almost undeniable -- to suppose that this difference has a basis in what is literally meant by those sentences.)

Suppose that, in (A) and (B), "two" and "the unique number n such that n=2 if there are continuous functions that cannot be differentiated at any point and such that n=1 otherwise" refer to the number two. In that case, it seems as though *both* (A) and (B) are doing the same thing. They are attributing a certain property to that number: the property of being a thing x such that John believes x to be greater than one. But if each of (A) and (B) attributes that property to that object, then they cannot differ in truth-value. So since they do differ in truth-value, we must suppose that they don't have ordinary referents. To sum up, consideration of Leibniz's Law, in connection with the behavior of expressions in indirect contexts, requires us to suppose that in such contexts expressions refer to their senses, not their ordinary referents.

So Frege holds the views previously stated *with* the qualification that, when an expression occurs in an oblique context, it undergoes a shift in reference: it refers to its sense.

Here is what I would like to argue. First, (SS) is wrong; and it is not needed to account for facts about cognitive significance. Second, Frege's views on indirect reference are wrong. Finally, those two views, in addition

to being wrong and unnecessary, are inconsistent with each other.

The Demolition of SS

Let us begin with a preliminary argument against (SS). Consider the sentence:

(*) "The inventor of bifocals snored".

Frege says that "the inventor of bifocals" is an expression that refers to Franklin. But Frege also says that Franklin himself is *not* a constituent of the proposition meant by (*). What is such a constituent is not Franklin, but is rather a "sense". A sense is a concept (in the non-psychological, platonic sense) – in this case, it would be the concept *inventor of bifocals* (or perhaps *unique inventor of bifocals*). So on Frege's view, the proposition meant by (*) has that concept or sense as a constituent.

For the sake of argument, let us suppose, with Frege, that the concept *inventor of bifocals* is a constituent of the proposition meant by "the inventor of bifocals snored". In that case, that proposition has the form:… *inventor of bifocals.*…So, presumably, that proposition *says something* about that concept. Supposing this is the case, what would that proposition say? We know that "the inventor of bifocals snored" is not going to be true if there is *no* inventor of bifocals. And we know that it won't be true of there are multiple inventors of bifocals. (In this context, let us leave aside issues relating to contextual salience.) So if that sentence is to be true there must be some x such that x uniquely invented bifocals. Of course, if x doesn't snore, then "the inventor of bifocals snored" will be false. So, in Russell's view, *if* the concept *inventor of bifocals* makes it into the proposition meant by "the inventor of bifocals snored", then that sentence encodes a proposition logically equivalent with: *for some x, x uniquely invented bifocals; moreover, x snored.* And if that proposition, or one logically equivalent thereto, is what is meant by the sentence in question, then it really does seem that "the inventor of bifocals" is a quantifier (a kind of existential quantifier). To sum up: *if* the concept *inventor if bifocals* is a constituent of *the proposition meant* by sentences of the form "…the inventor of bifocals…", then it seems that "the inventor of bifocals" is a quantifier, and not a singular term.

The other side of the story is that *if* "the inventor of bifocals" is a

singular term that refers to Franklin, then the concept *inventor of bifocals* does not make it into the proposition meant by "the inventor of bifocals snored". What *would* make it in would be Franklin himself: the sentence would encode the proposition *Franklin snored.* It would not, in that case, encode anything about bifocal-inventing.

Here is the long and short of it. Suppose that the proposition meant by (*) has the concept *inventor of bifocals* as a constituent. In that case, the proposition meant is identical with, or logically equivalent to:

(**) *For some x, x uniquely invented bifocals; moreover, x snored.*

But if (**) is the meaning of (*), then Russell's theory of descriptions is right: definite descriptions are quantifiers, not singular terms.

§ Admittedly, the argument just given is far from conclusive. There is a position we haven't yet explored. One could say that (*) is logically equivalent with (**), but that "the inventor of bifocals" is *still* a singular term.

More generally, *The phi has psi* is *logically equivalent* with *exactly one thing x has phi, and x also has psi*. But *the phi* is still a singular term. This is the position held by Anderson (1975). It is, I believe, the view that any Fregean *must* hold; and it is a wonder that it is not more widely advocated. For reasons we've seen, it is pretty clear that *if* the concept *inventor of bifocals* (as opposed to Franklin) is a constituent of the meaning of (*), *then* the proposition meant by (*) is logically equivalent with (**). Suppose that is right. In that case, *if* the definite description is a singular term, it immediately follows that (*) is logically equivalent with (**), but "the inventor of bifocals" is a singular term.

How would such a view be defended? The general conceit would be this. Semantics cannot be read off of logical equivalences. "1+1=2" and "triangles have three sides" are logically equivalent. But semantically they are as different as can be.

This position can be applied in a way that is more specific to the issues at hand. Given *only* that:

(A) *the phi has psi*

is logically equivalent with

(B) *something x uniquely has phi, and x also has psi*,

it by no means follows that *the phi* is a quantifier. After all,

(C) "Socrates is bald"

is logically equivalent with

(D) "Something x is uniquely identical with Socrates, and x is bald."

But "Socrates" is not a quantifier. In fact, *any* true sentence is logically equivalent with some sentence of the form "something x uniquely has phi and x has psi". For example, "turtles play golf" is logically equivalent with "there is some number n such that (n+1=2 and turtles play golf), and turtles are either square or not square". So semantics cannot always be read off of logical equivalences.

Nonetheless, I think *if* *the phi* is logically equivalent with its Russellian counter-part, then it really *is* just a quantifier. If (*) is logically equivalent with (**), then Russell's theory is right, and Frege's theory is wrong.

Consider (C). The *lack* of quantificational structure in (C) shows up when you negate it or otherwise embed it in a larger sentential context.

(CN1) "Socrates is *not* bald"

is in no way ambiguous. But if (C) had the same *meaning* as (D), then (C1) *would* be ambiguous between:

(CN2) "Something x is uniquely identical with Socrates, and x is *not* bald"

and

(CN3) "it is not the case that something is uniquely identical with Socrates, and x is bald".

Similarly,

(CM1) "Socrates is necessarily bald"

is *not* ambiguous. But if (C) had the same meaning as (D), then (CM1) would be ambiguous between

(CM2) "Something x is identical with Socrates, and necessarily: x is bald"

and

(CM3) "Necessarily: something x is identical with Socrates, x is bald"

(*)"Necessarily: Socrates is identical with Socrates"

is unambiguously true. But if "Socrates" were a quantifier of the kind in question, then (*) would be ambiguous between a true statement and a false one:

(**) Necessarily: something x is identical with Socrates, and x is identical with Socrates.

(***) Something x is identical with Socrates, and necessarily: x is identical with Socrates.

(**) is false; (***) is true.

So the fact that "Socrates" is *not* a quantifier is revealed by the ambiguity of sentences like (*) and (CM1), and the like.

Given this, suppose that

(#) *the phi has psi*

is logically equivalent with

(##) *something x uniquely has phi, and x has psi*.

In that case,

(#N) *necessarily: the phi has psi*.

would be ambiguous, as the necessity operator could be given different degrees of scope. For example,

(#N1) "the inventor of bifocals is necessarily identical with the inventor of bifocals"

becomes ambiguous between

(#N2) "necessarily: something x uniquely invented bifocals, and something y invented bifocals, and x=y"

and

(#N3) "something x uniquely invented bifocals, and something y uniquely invented bifocals, and necessarily: x=y".

(#N2) is false; (#N3) is true.

So the very things which show that "Socrates" is *not* a quantifier seem to show that *the phi*, as Anderson analyses it, *is* a quantifier. Where "Socrates is bald" can only be negated *one* way, *the phi has psi* can be negated in *two* ways. Where "necessarily Socrates is Socrates" is unambiguously true, *necessarily the phi is the phi* is only ambiguously true. And so on. So the very considerations that make "Socrates" *not* be a quantifier aren’t there to make *the phi* (as Anderson analyses it) not be a quantifier.

There is a similar line of thought to consider. Suppose that "Sukrat" is a proper name that is co-referential with "Socrates" (which, let us assume, is also a proper name):

($) "Socrates is bald. Socrates= Sukrat. Therefore Sukrat is bald."

($) is a genuine application of Leibniz's Law.

For the sake of argument, suppose that *Russell* (not Anderson) is right about definite descriptions. In that case:

($$) "The inventor of bifocals is bald. The inventor of bifocals=the first post-master general. Therefore the first post-master general is bald."

($$) is certainly a valid argument. But it is not a straightforward application of Leibniz's Law. As Neale (1990) points out, it is not an immediate application of Leibniz's Law, but rather a *theorem* (of the system of the PM) that:

($$$) The phi has psi. The phi=the chi. Therefore the chi has psi.

Of course, the system of the PM is a purely *extensional* one: it contains no troublesome – e.g. epistemic, modal -- operators. If it weren't, then ($$$) would simply be invalid.

In any case, ($$$) is surely not *identical* with Leibniz's Law.

It certainly seems that if

(#) *the phi has psi*

is logically equivalent with

(##) *something x uniquely has phi, and x has psi*,

then

($$) "The inventor of bifocals is bald; the inventor of bifocals=the first post-master general; therefore the first post-master general is bald"

is not a straightforward application of Leibniz's Law. The logic behind ($$) is more complicated than that behind ($), even though both are valid. And I don't see that the proof-theoretic differences between ($) and ($$) vanish if it is maintained that definite descriptions are singular terms (that are logically equivalent with Russellian quantifiers) – the proof of ($$) will *still* be different from that of ($). So if *the phi* is logically equivalent with its Russellian counter-part, the logic governing sentences containing it will be very close to, if not identical with, the logic governing Russellian quantifiers, and very different from the logic governing things that we *know* to be singular terms (e.g. "Socrates"). So it becomes unclear what it could y mean to say that *the phi* is logically equivalent with a Russellian quantifier but is *still* a singular term.

Leibniz's Law and the logic of identity statements

As we've noted, there are no *actual* exceptions to Leibniz's Law. So no legitimate semantic system can allow there to be *actual* violations of Leibniz's Law. Frege noted that there are apparent violations of Leibniz's Law (in modal, epistemic, etc. contexts). Partly for this reason, he said that, in such contexts, referring terms don't have their ordinary referents: this semantic shift prevents the occurrence of any actual violations of Leibniz's Law.

According to Neale (1990), if Russell is right, then

($$) "The inventor of bifocals is bald. The inventor of bifocals=the first post-master general. Therefore the first post-master general is bald"

is not an instance of Leibniz's Law at all. By the same token,

($$SB) "Smith believes that the inventor of bifocals is bald. The inventor of bifocals=the first post-master general. Therefore Smith believes that the first post-master general is bald"

is not a *violation* of Leibniz's Law.

If Anderson is right about definite descriptions, then (as we've

discussed) it seems that ($$) is not an instance of Leibniz's Law (though the proof of it *involves* that law). So by the same token, if Anderson's semantics is right, ($$SB) is not a *violation* of Leibniz's Law. So, on Anderson's semantics, a need to safeguard Leibniz's Law does *not* require us to suppose that epistemic (and modal…) contexts induce semantic shifts. After all, on Anderson's semantics, as on Russell's, ($$) is no more an instance of Leibniz's Law than ($$SB) is a violation of it.

Let us close this particular argument. Consider:

(i) *the phi has psi.*

Suppose that O is the unique phi. If Frege is right, then the concept *phi*, and not O, is what makes it into the proposition meant by (i). But if that is the case, then (i) is logically equivalent with:

(ii) *something x uniquely has phi, and x also has psi*.

It would also follow that:

(iii) *John believes that the phi has psi.*

would be ambiguous between two propositions, namely:

(iv) *something x uniquely has phi, and John believes that x has psi*.

and

(v) * John believes that something x uniquely has phi, and x has psi*.

So if, as Frege believes, the concept *phi* is what makes it into the proposition meant by sentences containing *the phi*, then *the phi* is acting like a quantifier and *it is doing the same thing in both direct and indirect contexts.* So Frege's view actually requires that we *not* suppose that *the phi* undergo some kind of semantic shift in oblique context. Frege's view that *the phi* *does* undergo such shifts is inconsistent with this view that it is the concept *phi*, and not O (the unique thing which has phi), that makes it into the proposition meant by a sentence of the form *the phi has psi*.

§ I believe that we have stumbled on a matter of some importance. Some terminology will help. As we've discussed, Frege holds the *sense* of "the inventor of bifocals" is what makes it into the proposition meant by: * the inventor of bifocals has phi*. Let us refer to this as Frege's Fundamental Assumption (FFA).

*(FFA) Suppose that O is the unique phi. When *the phi* is a part of a sentence, what makes it into the corresponding proposition is the sense of *the phi*, and not O itself.*

Here is a more formal way of putting (FFA). In virtue of having the form *…the phi…*, it *is* the case that a sentence encodes a proposition that has the *sense* of *the phi* as a constituent; and it is *not* the case that, in virtue of having that form, a sentence has O has a constituent.

What is not often realized is that, if FFA is right, then *the phi* is doing the very same thing in both direct *and* indirect contexts. Suppose FFA is right. Now consider the following sentence:

(ibfs) "the inventor of bifocals snored."

Because "the inventor of bifocals" occurs in (ibfs), the proposition thereby meant has the concept *inventor of bifocals* as a constituent. As a result, the proposition meant by (ibfs) is identical with, or at least logically equivalent to:

($ifbs_p$) something uniquely x invented bifocals, and x snored.

Now consider the sentence:

(jifbs) John believes that the inventor of bifocals snored."

Because "the inventor of bifocals" occurs in (jibfs), the proposition thereby meant has the concept *inventor of bifocals* as a constituent. If, as Frege maintains, the sense of the description makes it into the proposition meant by (jifbs), then that sentence is ambiguous. On one reading, the proposition thereby meant is identical, or at least logically equivalent, with:

(jifbs$_W$) Something uniquely x invented bifocals, and John believes that x snored.

On the other reading, the proposition meant by (jibfs) is identical with, or at least logically equivalent to:

(jifbs$_N$) John believes that something uniquely x invented bifocals, and x snored.

If FFA is right, then there is *one rule* for "the inventor of bifocals". That rule is:

(*) For any property psi, the proposition meant by *the inventor of bifocals snores* is (logically equivalent to): *something* uniquely x invented bifocals, and x has psi.

If "the inventor of bifocals" occurs in a molecular sentence, there will be two different ways to *apply* (*). There will be a wide-scope application and a narrow-scope application. But there will still be just *one* rule that applies to all contexts, whether direct or indirect.

Let us sum up this leg of the argument. Supose FFA is right. In that case, *the phi* does the same thing in indirect contexts that it does in direct contexts. In *every* case, what makes it into the proposition meant by *the phi has psi* is the *sense* of the definite description: and the sense doesn't change depending on the context. In direct contexts, the sense that makes it in is the concept *phi* (or perhaps *unique phi*); and the same is true in indirect contexts.

Now we can close the argument. If FFA is right, then in *indirect* contexts "the inventor of bifocals" doesn't refer to Franklin. Let me clarify this. It is essential to Frege's semantics that the following expressions refer to *different* propositions:

(1) "That the inventor of bifocals snored"
(2) "That the first post-master general snored".

If (1) and (2) really do refer to different propositions, then compositionality demands that we *not* see the definite descriptions therein as

referring to a *person.* We've just seen that, if FFA is right, then what the definite descriptions are doing in (1) and (2) is exactly what they are doing in direct contexts like (ibfs) and "the first post-master general snored". Since they are not referring to Franklin in (1) and (2), we must conclude they don't so refer in direct contexts either. Basically, if we accept FFA, then we must deny that definite descriptions are *ever* singular terms. Moreover, if we accept FFA, Frege's idea that definite descriptions undergo some kind of meaning-shift in indirect contexts becomes unnecessary: in fact, it becomes *de rigueur* not to hold that. For given FFA, *the phi* is doing the very same thing in both direct and indirect contexts.

There is another way of looking at this. Suppose that we (incoherently) accept *both* FFA *and* the view that *the phi* is a singular term. In that case, as Frege realized, compositionality demands that we see the occurrence "the inventor of bifocals" as referring to one thing, and its occurrence in (ibfs) as referring to something very different. And we must see it as referring to a third thing in: "it is true that the inventor of bifocals snored". That is counterintuitive. Indeed, as we will see, Davidson argued compellingly that it is absurd. As we will also see, that view is not needed to accommodate either compositionality or facts about cognitive significance.

§ This discussion leads us to another point. If *the phi* really were a referring term, then it seems to me that *the phi=the chi* would express an identity. If it expressed an identity, then it would be subject to the logical principles governing identities, the main one being Leibniz's Law. But when we want to derive *the chi has psi* from *the phi=the chi, and the phi has psi*, we have to invoke logical principles appropriate to *quantified* statements like *something x uniquely has phi, and x also has psi*. So if, as Anderson plausibly maintains, *the phi has psi* is logically equivalent with its Russellian paraphrase, then from a proof-theoretic standpoint, *the phi=the chi* does *not* (merely) express an identity; from a proof-theoretic standpoint, it seems to express a complicated existence claim; from that standpoint it is much more like its Russellian paraphrase than it is like anything of the form *a=b* (where a and b are proper names). And this again raises the question how *the phi* could be logically equivalent with a Russellian quantifier without actually being one.

We have to distinguish identities from identity-statements. Identities either hold necessarily or they don't hold at all. Nothing can be contingently

self-identical, and there is no other kind of identity besides self-identity; so identities hold necessarily (if at all). But the truth expressed by a so-called identity-*statement*, e.g. "the inventor of bifocals=the first post-master general", may be contingent. To the extent that a statement expresses a *contingent* truth, it is not expressing an *identity*, as we just saw. It is certainly arguable that *part* of what is expressed by "the inventor of bifocals=the first post-master general" is an identity. But what that sentence says is (on a natural readings of it) contingent, and is therefore not (merely) an identity.

It seems that in so far as E1 and E2 really *co-refer*, then *E1=E2* should express an identity and, therefore, a necessary truth. , I will now try to argue that *E1=E2* can express a contingent truth only to the extent it does *not* express an identity and, therefore, only to the extent that they are not co-referring terms. If E1 refers to x, then surely *E1 has phi* ascribes phi to x.[3] By the same principle, if E2 refers to x, then *E2 has phi* should ascribe phi to x. One property that can be ascribed to a thing x is the property of being identical with x. So, in so far as E1 is an expression that refers to x, *E1=E1* should ascribe the property being identical with x to x. And in so far as E2 is an expression that refers to x, *E1=E2* should ascribe that same property to x. So in so far as E1 and E2 are both expressions that refer to x, *E1=E2* must express a necessary truth. By the same token, in so far as *E1=E2* expresses a contingent truth, then either E1 or E2 (or both) is doing something *other* than referring to x.

Let us sum up. For the sake of argument, suppose that, as Frege maintains, it is not Franklin himself, but rather the sense of "the inventor of bifocals" is what makes it into the proposition meant by

(*) *the inventor of bifocals has psi*

In that case, the proposition meant by that sentence is logically equivalent, and possibly even identical, with:

(**) *something x uniquely invented bifocals, and x has psi.*

But if (**) is the meaning of (*), then the logic governing definite descriptions like "the inventor of bifocals" and "the first post-master general" is very different from the logic governing "Franklin" and any other proper name referring to Franklin; and, at the same time, that logic *will* be a lot like

that governing quantifiers. So *if*, as Frege maintains, the sense of "the inventor of bifocals" makes it into the proposition meant by sentences containing that expression, then that expression is a quantifier. Considerations of logic, especially Leibniz's Law, compel us to hold this.

Predication versus Reference

Of course, considerations of Leibniz's law were not Frege's only motivation for his view that *the phi* refers to its sense in oblique contexts. Considerations of compositionality were his main reason. But considerations of compositionality *don't* warrant that view, contrary to what is generally held. This becomes apparent when we discuss the difference between predication and reference.

In this context, let us assume for the sake of argument that (FFA) is correct. After all, if FFA is false, then so *ipso facto* is Frege's semantics.

Given this, consider the sentence:

(&) "Franklin=the inventor of bifocals"

Focus on the expression on the right side of the "=". What does that expression introduce into the proposition meant by (&)? (Remember that we are operating within Frege's view that, when *the phi* occurs in a sentence, what makes it into the proposition in question is the *sense*, not the referent, of that expression.) Answer: the sense of "the inventor of bifocals" – the concept *(unique) inventor of bifocals.*

Clearly (&) is doing much more than ascribing the property of being identical with Franklin to Franklin. (Actually, it isn't clear to me that even *part* of what (&) does is ascribe self-identity to Franklin.) The definite description has what Strawson described as a "predicational" function: the property of being a unique inventor of bifocals is predicated of Franklin. Perhaps we can see the definite description as a singular term that refers to Franklin. But plainly it is also functioning as a device of predication. What makes (&) non-trivial is precisely that it conveys the message that Franklin invented bifocals. So (&) is non-trivial because the definite description picks out, or is part of an expression that picks out, a *property*.

In so far as the definite description is a device of predication, it seems natural to see it as picking out a *property*, and not (merely) a person. It

is a little hard to see how *x=the inventor of bifocals* could ascribe the property of being a unique bifocal inventor to x unless the "=the inventor of bifocals" picked out that property; and it's a little hard to see how that expression could pick out that property if the definite description did not, in its turn, pick out at least some of the properties constitutive of that property. So I am not completely out on a limb when I say that, in functioning predicatively, the definite description in (&) can be seen as (inter alia) picking out a property (or multiple properties). This is not to deny that it *does* pick out a person. But it apparently also has a predicative function, and thus picks out a property. So the extent that definite descriptions function predicatively, they do not *merely* refer to individuals. That is probably *part* of what they do; but they also, so it appears, pick out properties.

If this is right, then it is highly misleading to say that "the first post-master general" *co-refers* with "the inventor of bifocals". At most we could say that they *partially* co-refer: each refers to Franklin, among other things. But since each obviously picks out a *property* as well – a different property in each case – it is not possible to say that *on the whole* they co-refer. So to that extent, they *diverge* in reference.

So if we accept Frege's own semantics, then we must say that "the first post-master general" and "the inventor of bifocals" *do not* co-refer. If we accept Frege's views about what makes it into the proposition meant by "the inventor of bifocals/the first post-master general snored", then we are compelled to *reject* Frege's view that the definite descriptions in question co-refer.

There is more to say in this connection. (Let us continue to suppose that (FFA) is right.) It is not just in the context of identity-statements that definite descriptions have a predicative function. It seems that they *always* have such a function. Consider our paradigm:

(IBS) "the inventor of bifocals snored".

This certainly says (in *some* reasonable sense of the word "says") that *somebody* or other invented bifocals, and that any such person snored. So the definite description *is* attributing a property to someone. In any case, in so far as (IBS) differs in cognitive value from

(FS) "Franklin snored",

it is because (IBS) is to the effect that someone or other invented bifocals, whereas (FS) does not. So in so far as those sentences have different cognitive values, it is because the definite description in (IBS) ascribes a property to someone or other – it is because the definite description is functioning predicatively. Given that it is so functioning, it is not absurd to see it as picking out a property (that of being a unique bi-focal inventor). By exactly similar reasoning, it is not implausible to see the definite description in:

(PGS) "the first post-master general snored"

as functioning predicatively and thus as *inter alia* picking out a property. Of course, the property picked out would be *different* from that picked out by "the inventor of bifocals". So even though there is overwhelming evidence to support the view that "the inventor of bifocals" and "the first post-master general" both refer to Franklin, there is also evidence to suggest that they also refer to other things, different things in each case. So even though they both refer to Franklin, neither *just* refers to Franklin, and *on the whole* they do not co-refer: they *overlap* in reference. So the steps in the Slingshot that depend on the idea that they (and other such pairs) "co-refer" involve, I believe, a muddle. Saying that those expressions co-refer is like saying that "Bill and Ted" and "Bill and Steve" co-refer.

So far as I can tell, the only way to block this conclusion is to say that "the inventor of bifocals" can function predicatively *without* referring to any property. In defense of such a position, it might be said that it "expresses" or perhaps "connotes" a property without actually *referring* to it. But this move puts a lot of weight on what, arguably, is a mere matter of nomenclature. When you say that "the inventor of bifocals" expresses or connotes the property of being a unique bifocal inventor, what you are saying is that *the inventor of bifocals has phi* says, or at least implies, that somebody or other uniquely invented bifocals, and that any such person has phi.[4]

So it is obviously part of the *semantics* of *the inventor of bifocals has phi* that it implies, if it doesn't outright say, that somebody or other invented bifocals: we don't need to invoke any logic to establish this connection. So it is part of the semantics of "the inventor of bifocals" that

the inventor of bifocals has phi ascribes the property of being a bifocal inventor to someone. Given this, it is a little hard to see how that expression could *fail* to pick out that property. We can use the term "express" to describe this picking out. But a picking out – a referring – is what we are dealing with. This is not to deny that "the inventor of bifocals" also picks out Franklin – my own view is that it *does* pick him out. But it obviously doesn't *only* do that; and in so far as it doesn't only do that, it seems to have a predicative function, and therefore would appear to pick out properties. For exactly similar reasons, it is not implausible to see "the first post-master general" as picking out, not *only* Franklin, but also the property of being a first-post-master general – this being a property *not* picked out by "the inventor of bifocals".

Let us sum up. Suppose (FFA) is right. In that case, it follows, ironically, that those expressions *don't* co-refer: they end up referring to different properties. (Montague would say that they refer to different *functions.* But we'll see that these views collapse into each other.) So we would be forced to *deny* Frege's other view – his belief that they are co-referring singular terms. So Frege's system is self-contradictory.

Further, more importantly, this shows that compositionality does *not* require us to see sentences as referring to truth-values *even if* we grant Frege's fundamental assumption. Even if we grant (FFA), when we have two expressions that have different senses but allegedly co-refer, it turns out that they refer to *different* properties.

Some other inconsistencies in the Fregean View

According to Frege, "the king of France is bald" *presupposes* but does not itself affirm, the existence of a king of France. This means that *if* there is no king of France, then that sentence is without truth-value.

This is certainly a reasonable position -- one I myself will defend. But it is not consistent with Frege's Fundamental Assumption. Suppose FFA is right. In that case, as we've seen, *the phi has psi* is logically equivalent with: "something x uniquely has phi, and x also has psi.* In particular,

(i) "The king of France is bald"

is logically equivalent with:

(ii) "Something x is uniquely a king of France, and x is also bald."

If there is no king of France, then (ii) is false. The absence of a king of France doesn't make (ii) without truth-value at all. The same is true of *any* sentence S encoding a proposition logically equivalent with: *something x is uniquely* a king of France, *and x is also bald.* So any such sentence does *not* presuppose the existence of a king of France, but actually affirms it. So Frege is guilty of inconsistency.

This criticism is not an *ad hominem* attack on Frege's particular system; it has more general consequences for semantics. If FFA is right, then *the phi* *asserts* the existence of a unique phi. But in that case, *the phi has psi* is simply *false* if there is no (unique) phi, meaning that *the phi has psi* does *not* presuppose existence of a phi. In that case, *the phi* does not *refer* at all: where there is reference, there is a presupposition of existence, not an affirmation of it. If you say "that man [pointing to Fred] is a scoundrel", you are not *affirming* the existence of a contextually salient man; you are presupposing it. If you say "George's literary career is really coming along", you are not affirming, but presupposing, George's existence. Given any clear case of reference, we see that referring presupposes existence, and doesn't assert it. So if FFA is right, then *the phi* is *not* a singular term, contrary to what Frege held. If we are to hold onto FFA, then we must jettison the idea that *the phi* is a singular term; we must see it as affirming existence: we must see *the phi has psi* as meaning: *something x uniquely has phi, and x also has psi.* So we must see *the phi* as being a quantifier – we must accept the Theory of Descriptions.

I wish to make my own position clear. I think that *if* FFA is right, *then* one must accept the Theory of Descriptions. But I think that both FFA and the Theory of Descriptions. I believe they both confuse semantics with *pre*-semantics and, what is closely related, type-semantics with token-semantics.

Is "reference" an ambiguous term? Does Anderson's semantics make it ambiguous?

Anderson's position is this. *The phi* is a singular term but *the phi has psi* is logically equivalent with its Russellian paraphrase. Strictly for the sake of argument, let us suppose that Anderson is right about this.

Let us now *invent* an expression such that, by our stipulation, what Russell said about *the phi* is actually true of that other expression. Let *the# phi* be an expression such that everything Russell said about *the phi* is true of *the# phi*.

Now consider:

(i) "The inventor of bifocals snores", and
(ii) "The# inventor of bifocals snores".

The relation that Franklin's snoring (or not snoring) has to the truth-value of the one sentence is identical with the relation that it has to the truth-value of the other. E sentence expresses a proposition P that exists, and can be true or false, in worlds where Franklin doesn't exist; each expresses a proposition that is true in worlds where Franklin uniquely invents bifocals and does snore, and that is false in worlds where Franklin uniquely invents bifocals and does not snore. In general, supposing that O is the unique phi, the relevance of facts about O to the truth of *…the phi…* will be identical with the relevance of such facts to the truth of *....the# phi…* So it's a little hard to see how *the phi*, as Anderson analyses it, is significantly different from a Russellian quantifier.

Admittedly, this by itself is not probative. Let us stipulate that "Socrates#" is to be defined as follows:

Socrates# has phi means: something x is uniquely identical with Socrates, and x has phi.

So "Socrates#" is a quantifier, even though "Socrates" is not. Now the relevance which Socrates' snoring (or not snoring) has to:

(&) "Socrates snored"

is identical with that which it has to:

(&&) "Socrates# snored".

But "Socrates#" is a quantifier, whereas "Socrates" is not. So, to close the argument, the point just made about Franklin's relevance to (i) and (ii) doesn't settle whether "the phi" (as Anderson analyses it) is a quantifier or not.

What are we to say in response to this? First of all, notice that, if a sentence comprises an occurrence of "Socrates#", that sentence has for its meaning a proposition that has Socrates himself as a constituent. *As a whole*, "Socrates#" is a quantifier, and not a singular term referring to Socrates. But, for the reason just described, it is not out of the question to see Socrates as *part* of what "Socrates#" denotes.

It seems to me (though here I am entering into the vagaries of psychology) that a person can *understand* "Socrates snored", and know that it is true, without knowing that:

(%) somebody x uniquely was identical with Socrates, and x snored.

Permit me a bit of science-fiction. Suppose that Smith has strange cognitive deficits whereby he can grasp singular propositions (like *Franklin smokes*), but not general propositions (like *all men smoke*). It is conceivable that animals are in such a position. (I am fairly certain that squirrels grasp singular propositions, like *that dog is eating my acorn*, but I have some doubts about whether they can grasp generalizations of such propositions.) Smith would, I think, be able to understand (&), but not (&&). So the semantic or linguistic competence needed to understand – to grasp the meaning of -- (&) is not sufficient for a grasp of (&&). Somebody who understands (&) needs to call on extra-semantic *logical* acumen to see that it entails (%); linguistic competence is not enough.

Let me put this another way. "Socrates snored" is true exactly if Socrates snored. This is a matter of semantics: in virtue of having the right bit of linguistic or semantic competence, one knows that "Socrates snored" is true iff Socrates snored. For obvious reasons, "Socrates snored" is true exactly if somebody x is uniquely identical with Socrates, and x snored. But this is a matter of *logic*, not (merely) of semantics. The semantic rule for

"Socrates snores" is simply: "Socrates snored" means *Socrates snored* (or is true iff Socrates snored). The semantic rule doesn't involve any mention of quantifiers or identity. So the bit of *semantic* competence that is needed to understand "Socrates snored" is not, by itself, sufficient for knowing that "Socrates snored" is true iff something x is uniquely identical with Socrates, and x snored. This gives us some reason to think that what "Socrates snored" *means* is not that existence claim, or (by a similar argument) any other quantified statement, even though what it means is logically equivalent with that claim; and, therefore, that "Socrates" isn't a quantifier (even though it is logically equivalent with one).

A slight digression may help. "John has two apples" is true exactly if

@) either there are square circles or the number n such that John has n apples is equal to the number of square roots of a positive number.

But the semantic rule for "John has two apples" is just – that sentence means (or is true iff) John has two apples. There is no mention of square roots or the property of being a positive integer. So the connection between "John has two apples" and @ is not (merely) semantic; extra-semantic (mathematical) principles have to be brought in.

It seems to me that (&) and (&&) don't differ from each other in the way in which "Socrates snored" and "Socrates# snored" differ. As a matter of semantics, (&) is true exactly if something x uniquely invented bifocals and x snores. And the same is true of (&&). The equivalence between (&) and (&&) seems not to be comparable to the equivalence between @ and "John has two apples" or to that between "Socrates snored" and "Socrates# snored". The latter equivalences are mediated by extra-semantic, logical principles. The first equivalence seems to be purely semantic.

The Possible Ambiguity of the term "Reference" and the Distinction between Rigid and Non-rigid Designation

For the sake of argument, let us continue to suppose that Anderson's analysis of definite descriptions is correct.

In that case, the relation that holds between "Franklin" and Franklin

is very different from that which holds between "the inventor of bifocals" and Franklin. "Franklin snores" is true exactly if Franklin snores. It is irrelevant if Smith snores or if Jones doesn't snore. All of this is obviously bound up with the fact that "Franklin" refers to Franklin. If "Franklin snores" were made true by Smith's snoring, then (holding constant the other semantic rules of English), "Franklin" would refer to Smith.

Now "the first post-master general snores" is true in worlds where Franklin doesn't exist, and it is true in worlds where he does exist but doesn't snore. The most that can be said is that *if* Franklin is the inventor of bifocals, *then* "the first post-master general snores" is true iff Franklin snores. But the same can be said of *anything.* It is true that *if* Bill Clinton is the inventor of bifocals, then that sentence is true iff Clinton snores. The *semantic* relation between "the first post-master general" and Franklin is extremely loose.

It is pretty clear that, in the sentence "some one person invented bifocals", there is nothing that *refers* to Franklin.[5] It seems that the very considerations which establish this also establish that nothing refers to Franklin in the sentence "something x uniquely invented bifocals, and x snored". And it's a little hard to see why those same considerations wouldn't establish the same thing in the case of "the inventor of bifocals snored".

In any case, it is clear that the relation between "Franklin" and Franklin is not much like the relation of "the inventor of bifocals" to Franklin. Given this, the question can legitimately be raised whether we are rendering the notion of reference ambiguous if we say that *both* "Franklin" and "the inventor of bifocals" are terms that "refer" to Franklin.

Here it seems reasonable to reply by saying (as Anderson does[6]): "all we're dealing with here is the distinction between rigid and non-rigid reference; 'Franklin' rigidly, and "the inventor of bifocals" non-rigidly, refers to Franklin."

There are two points to make in response. First, if that response is right, then there are very good reasons to think that "non-rigid reference" isn't reference at all. Supposing that *the phi has psi* is logically equivalent with (one of its) Russellian paraphrase(s), we've seen that it is very hard to see why *the phi* can be said to be a singular term at all. So if we insist on saying that *the phi* "non-rigidly refers", then "non-rigid reference" is just another way of saying "non-reference".

§ But there is another difficulty with the view in question. To say that

"the inventor of bifocals" "non-rigidly" refers to Franklin is an entirely different thing from saying that it is logically equivalent with its Russellian counterpart. In other words, given only that "the inventor of bifocals" non-rigidly refers to Franklin, it could *still* be the case that "the inventor of bifocals snores" encodes the proposition: *Franklin snores.* Let me clarify.

Demonstratives are paradigm cases of non-rigid designators. "You" refers to different people on different occasions of use. There is no one thing x such that any given token of *you are phi* means that x has phi. But for any given token of *you have phi*, there is some x such that that token means *x has phi.*

Let us consider a case that is a little closer to home. Let us invent an article "the#" to be defined as follows. Let w be an arbitrary world, and suppose that, in w, O is the unique thing having phi. In that case, a token of *the# phi has psi* in w encodes the proposition: *O has psi.* So *the# phi* is a demonstrative: it picks out the thing which, in the world of utterance, has phi; and *the# phi has psi* encodes a singular proposition to the effect that that thing has psi (just as a token of *you have psi*, addressed to Fred, encodes a singular proposition to the effect that Fred has psi). So "the# inventor of bifocals snores" encodes the proposition: *Franklin snored.* (This is what Strawson said about "the".)

Obviously "the# inventor of bifocals" is not a rigid designator. In w2, it refers to Jefferson. In w3, it refers to Hartry Field. But, in any world w, a token of "the# inventor of bifocals snored" means *O snored* (where O is the unique inventor of bifocals). So even if an expression E *non-rigidly* refers to Franklin, it may still be the case that *E snored* means: *Franklin snored.* So the semantic differences between "Franklin" and "the inventor of bifocals" – that fact that "Franklin snored" means *Franklin snored*, whereas "the inventor of bifocals snored" means something that can be true in worlds where Franklin doesn't exist or does exist but doesn't snore -- cannot be dealt with *merely* by saying that "Franklin" rigidly denotes Franklin, whereas "the inventor of bifocals" non-rigidly him.

§ Now let us apply the points just made to the Slingshot. Suppose that when "the inventor of bifocals" occurs in a sentence (or sentence-token, to be precise), the resulting proposition has the sense of that expression as a constituent. In other words, suppose that FFA is right. In that case, "the

inventor of bifocals snored" means *something x uniquely invented bifocals, and x snored.* So if FFA is right, then "the inventor of bifocals" ends up being a quantifier, and Russell's theory of descriptions is right. We know that if the theory of descriptions is right, then every form of the Slingshot is blocked. So each form of the Slingshot requires that FFA be wrong: when *the phi* occurs in a sentence, what makes it into the proposition is *not* the sense of that expression, but only the thing (if any) referred to. So if O is the unique phi, then the proposition meant by *the phi snores* is simply *O snores.* And (supposing that O also uniquely has property chi) the proposition meant by *the phi is identical with the chi*, is simply *O=O.* This can be encapsulated in the slogan: if definite descriptions are singular terms, they are directly referential.

(The sense of a definite description would still have an important semantic role: but that role would simply be to *pick out* the thing which makes it into the proposition in question, and that sense would not itself make it in there. Even relegated to that role, that sense will still do the relevant *cognitive* work; it will still suffice to make "the inventor of bifocals snored" and "the first post-master general snored" *convey* radically different propositions. This is because, as we will see, much of what is communicated by an utterance is *pre*-semantic.)

Given this, let us look at some famous Slingshots.

Church 1943[7]

Assume both (LL) and (CR). Let S and S* be any two non-analytic sentence-tokens that have the same truth-value.

(1) S

(2) The class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x).

(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).

(4) S*.

(1) and (2) are logically equivalent. So (1) and (2) co-refer. (3) is what results when a referring term in (2) is replaced with a co-referring term. So (2) and (3) co-refer. (3) and (4) are logically equivalent. So they co-refer.

So (1) and (4) co-refer. Any two sentence-tokens, alike in truth-value, co-refer.

But this argument fails. Let Omni be the class of all self-identical objects. If that argument is to go through, we must read the definite descriptions as singular terms. But remember that, if *the phi* is a singular term referring to O, then *the phi has psi* encodes the proposition: *O has psi.* So if we read the definite descriptions as singular terms, we have:

(1_R) S
(2_R) Omni=Omni.
(3_R) Omni=Omni.
(4_R) S*.

There is no logical equivalence between the first two steps, or the last two. The argument fails.

Let us now consider Church 1956. Let us assume:

(CR) Replacing expressions with co-referring expressions cannot result in a change of *reference.*

(SR) Expressions that are synonymous (or "very nearly so") co-refer.

According to Church (1956), if we grant these principles, then we can prove that sentences refer to truth-values:

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is the author of Waverly.
(iii) Sir Walter Scott is the man who wrote twenty-nine Waverly novels altogether.
(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) The number of counties in Utah is twenty-nine.

By CR, (i) and (ii) and (iii) must all refer to the same thing.

(iii) and (iv) are, in Church’s view, synonymous (or "very nearly so").

Nearly synonymous expressions must co-refer. So (iii) and (iv) co-refer.

By CR, (v) refers to the same thing as (iv).

It follows that (i) and (v) refer to the same thing. They also have the same truth-value. Indeed, the one significant thing that (i) and (v) have in common is that they are both true.

CR preserves reference. The one significant property of sentences that CR preserves is truth-value. So sentence-reference is sentence truth-value.

What is true of (i) and (v) is true (mutatis mutandis) of *any* two true sentences. Let S and S* be any two true sentences you choose. Given CR, plus the innocuous principle that synonymous expressions co-refer, a little bit of syntax-chopping shows that S and S* have the same referent.

Thus any two true sentences refer to the same thing: the property of truth.

An exactly similar argument shows that any two false sentences refer to the same thing: the property of falsity.

But given what we've said about reference, it is clear that Church's argument fails. Church assumes that "Scott" and "the author of Waverly" both refer to Scott. Suppose he is right. In that case, for reasons we've seen, if O is the author of Waverly, then what is meant by *the author of Waverly has chi* is: *O has chi.* So if, as Church assumes, "the author of Waverly" is a singular term referring to Scott:

(i) "Scott was wise"

and

(ii) "The author of Waverly was wise"

ought to mean the same thing. They both ought to mean:

(iii) *Scott was wise.*

So lines (i) and (ii) must *both* mean:

(i_R) *Scott is Scott.*

Church assumes that "twenty-nine" refers to the same thing as "The number of counties in Utah" and also "the number n, such that Sir Walter Scott wrote n Waverly novels". He assumes they all refer to the number twenty-nine.

Let us suppose he is right. In that case, by reasoning exactly similar to that just given, lines (iii)-(v) must all mean:

(iii_R) The number twenty-nine is identical with the number twenty-nine.

Thus (i)-(v) are synonymous with:

(i_R) *Scott is Scott.*
(ii_R) *Scott is Scott.*
(iii_R) *Scott is Scott.*
(iv_R) The number twenty-nine is identical with the number twenty-nine.
(v_R) The number twenty-nine is identical with the number twenty-nine.

(iii_R) and (iv_R) are not synonymous. The argument fails.

A Salmon-Kripke reply

I must consider a response to my argument (due to Nathan Salmon):

Here is a response to your argument. Never mind English definite descriptions. Instead we introduce an artificial device "ze" that works according to the following rule: *ze phi*[8] refers to the unique individual that satisfies (has) phi, if there is such an individual, and fails to refer otherwise. In that case "Scott" and "ze author of Waverly" are co-referential. Now replace "the" through Church’s (or Frege’s) argument with "ze". The resulting argument is immune to what you’ve said.

Let us see how well this proposal fairs.

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is ze author of Waverly.
(iii) Sir Walter Scott is ze man who wrote twenty-nine Waverly novels altogether.
(iv) Ze number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) Ze number of counties in Utah is twenty-nine.

Of course, "ze author of Waverly" contributes the very same thing as "Scott". They both contribute Scott. So the proposition encoded in (ii) is identical with the proposition encoded in (i).

"Ze man who wrote twenty-nine Waverly novels altogether" contributes Scott. So (iii) encodes the same proposition (ii) and as (i).

"Ze number n, such that Sir Walter Scott wrote n Waverly novels" refers to/contributes the number twenty-nine. So (iv) encodes the proposition: twenty-nine is twenty-nine.

"Ze number of counties in Utah" refers to/contributes the number twenty-nine. So (v) encodes the proposition: twenty-nine is twenty-nine.

So when we rewrite Church's argument, replacing "the" with "ze", and consider the propositions encoded in the resulting sentences, the result is this:

(i) Scott is Scott.
(ii) Scott is Scott.
(iii) Scott is Scott.
(iv) Twenty-nine is twenty-nine.
(v) Twenty-nine is twenty-nine.

(iii) and (iv) are not synonymous, or "very nearly so". So the argument fails.

Let us refer to expressions of the form "ze phi" as "zefinite descriptions".

By our stipulation, "ze unique x such that 1+1=2 and x=Plato" semantically contributes Plato and Plato alone. So "Plato snores" is synonymous with "ze unique x such that (1+1=2 and x=Plato) snores" and with "ze unique x such that (arithmetic is incomplete and x=Plato) smokes." This blocks Frege's argument.

In general, replacing a referring term with a co-referring zefinite-description preserves meaning. Replacing definite descriptions with zefinite descriptions will not help any slingshot; it will actually ensure that meaning *is* preserved and will prevent the divergence of meaning from truth-value that slingshots depend on.

Davidson 1967[9]

Assume (LL) and also that "logically equivalent expressions co-refer". Pick any two true sentence-tokens. Pick, say, "grass is green" and "whales are mammals". (I)-(IV) are meant to be sentence-tokens, not propositions:

(I) grass is green.

(II) the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x).

(III) (x=x and whales are mammals) is identical with the class of all things x such that (x=x).

(IV) whales are mammals.

(I) and (II) must co-refer. For given (LL), "grass is green" must refer to the same thing as "the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)". So, given (CR), (I) and (II) co-refer.

(II) and (III) co-refer for the same reason. And (III) and (IV) co-refer for the same reason. So (I) and (IV) co-refer. Any two true sentences co-refer.

Given this, suppose that sentences refer to the facts or states of

affairs that they indicate. Suppose, for example, "grass is green" refers to the fact that grass is green. In that case, it follows that "grass is green" refers to the same fact as "whales are mammals". It follows that any two true sentences refer to the *same* fact. Davidson bases far-reaching conclusions about semantics on the basis of this fact (no pun intended).

But no such conclusions are warranted, at least not by this argument. Again, let Omni be the class of self-identical things. If Davidson's argument is to go through, definite descriptions must be treated as singular terms. In that case, given what we've seen about reference, what is meant by * the class of all things x such that (x=x and grass is green) has phi* is: *Omni has phi*. In particular, what is meant by (II) is simply: *Omni=Omni*, the same being true of (III). So if we treat definite descriptions as singular terms, and perspicuously state what is going on, we are left with this:

(I_R) grass is green.
(II_R) *Omni=Omni*
(III_R) *Omni=Omni*
(IV_R) whales are mammals.

But (I) and (II_R) aren't logically equivalent, and Davidson's argument fails.

In another publication, Davidson uses a similar argument to show that there is only one fact. (On the basis of this supposed truth, Davidson advances some astonishing views about the nature of representation – he says there are no representations) Here is a reconstruction of his argument. Consider the following four expressions:

(I_F) "The fact that grass is green."
(II_F) "The fact that the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)."
(III_F) "The fact that (x=x and whales are mammals) is identical with the class of all things x such that (x=x)."
(IV_F) "The fact that whales are mammals."

(I_F) and (II_F) must co-refer. For given (LL), "grass is green" must refer to the same thing as "the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)". So, given (CR), (I_F) and (II_F) co-refer.

(II_F) and (III_F) co-refer for the same reason. And (III_F) and (IV_F) co-refer for the same reason. So (I_F) and (IV_F) co-refer. Any two expression of the form "the fact that…" co-refer. There is only one fact.

An exactly similar argument proves that there is only one true proposition. Simply replace each occurrence of "the fact that" in the foregoing argument. It is then established that "the proposition that grass is green" co-refers with ""the proposition that whales are mammals", and with any other expression denoting a true proposition.

Given what we've seen about reference, this argument of Davidson's is simply fallacious. Suppose that, as Davidson's argument requires, we treat definite descriptions as singular terms. In that case, for any property phi, *the class of all things x such that (x=x and grass is green) has phi* means simply: *Omni has phi.* So when we perspicuously re-write (I_F)- (IV_F), we are left with:

(I_F) The fact that grass is green.
(II_F) The fact that Omni=Omni
(III_F) The fact that Omni=Omni.
(IV_F) The fact that whales are mammals.

Since "Omni=Omni" is not logically equivalent with "grass is green" or with "whales are mammals", the first and last steps in the argument fail. Exactly similar considerations show the invalidity of the corresponding argument regarding propositions.

A Quinean Slingshot

Not too many people believe that there is only one fact or one proposition. Nobody believes that there is only one physical object.

But many people believe, or used to believe, that *modal* operators,

like "necessarily" and "possibly", are meaningless or result in systematic falsity. This position is associated with hard-line Humean empiricism.

Russell[10] held that "necessarily" and "possibly" are meaningless or superfluous. He said, first, that necessity and possibility are not properties of propositions, but of propositional functions. And he said that for a propositional function to be "necessarily" true is simply for it to be true for all values of its variables. Of course, he couldn't say that a propositional function is "necessary" if it is true for all *possible* values of its variables; that would make his analysis circular, since "possible" and "necessary" are interdefinable.

So Russell took the counter-intuitive position of saying that for a propositional function to be necessary is for it to be true for all *actual* values of its variables. So, for Russell, what is expressed by "for any time t, Plato did not smoke a cigarette at t" is necessary, since it is true for all actual values of its variables. But this analysis of necessity doesn't correspond even remotely to our modal judgments.

A true proposition is vacuously true for all values of its free-variables. So Russell has said, in effect, that any two true propositions are necessary, and any two false ones are impossible.

Plainly, Russell has not analyzed the concept of necessity; he has banished it from discourse.

Quine used the Slingshot to reinforce this kind of antipathy towards modality. Quine had a special reason for wanting to show that modal notions were corrupt. In "Two Dogmas of Empiricism", he argued that there are no analytic sentences. In effect, he argued that there are no necessary propositions. Quine didn't distinguish epistemic from metaphysical necessity; he thought that, if there was any necessity at all, it was the kind characteristic of "1+1=2"; he didn't recognize the kind (supposedly) characteristic of "water is H_2O".

So Quine needed there to be no epistemically necessary propositions; for this thesis was a cornerstone of his views on language and scientific explanation. He thought, presumably, that his Slingshot provided independent corroboration for this thesis.

It is worth showing why Quine's Slingshot fails. In so doing, we will curb the modal nihilism, still alive today, characteristic of empiricism.

Some sentence-connectives are "extensional". The truth of "snow is white and grass is green" is wholly a function of the truth-values of the

sentences falling within the scope of the "and". So "and" is extensional. Some sentences very much appear *not* to be extensional. The truth of "Khrushchev resigned after Kennedy died" is *not* a function of the truth-values of the constituent sentences ("Kennedy died", "Khrushchev resigned"). Both those sentences are true; but that is compatible with "Khrushchev resigned after Kennedy died" being either true or false.

Orthography doesn't always make it clear what the *scope* of a sentence connective is. In the sentence "Grass is green and snow is white", the *scope* of the "and" is really the (ordered) pair <grass is green, snow is white>.[11] English-notation doesn't make this fact clear. (So-called "Polish" logical notation does make it clear.) The "and" can be thought of as a function that assigns truth to a sequence of expressions exactly if both of them denote true propositions. But what is important is that *both* sentences fall within the scope of the "and".

The same thing is true of "after". In the sentence "Khrushchev resigned after Kennedy died", *both* constituent sentences fall within the scope of the "after", even though only one of those sentences is to the right of it.

Two more pieces of terminology must be defined. Let us say that a connective is PSLE[12] iff there is no change in truth value as a result of replacing a sentence falling in its scope with a logically equivalent sentence. A clear example of such a connective is "necessarily".

Finally, let us say that a sentence-connective "induces opacity" iff sentences falling within its scope do not have their ordinary meaning. "And" does not induce opacity; for "snow is white" has the same meaning in "snow is white and grass is green" that it has on its own. I cannot give an uncontroversial example of a connective that induces opacity. I myself don't think that *any* sentence-connectives induce opacity. That is actually one of the things I am trying to show. But Frege thought that epistemic connectives – e.g. "John believes that" -- induced opacity.

A Slingshot-style argument can be constructed that shows that at least *one* of the following two propositions is true:

(TF) All PSLE sentence-connectives are truth-functional.
(OP) All PSLE sentence-connectives induce opacity.

Let "Phi" be an arbitrary PSLE connective. And let S and S* be any two sentences that have the same truth-value. And let us refer to the

forthcoming argument as *TF*

(1) Phi (S)".
(2) Phi (the class of all things x such that (x is identical with x and S) is identical with the class of all things x such that (x is identical with x)).
(3) Phi the class of all things x such that (x is identical with x and S*) is identical with the class of all things x such that (x is identical with x)).
(4) Phi (S*).

Given that "Phi" is PSLE, we clearly have safe passage from (i)-(iv). It seems to follow that *any* PSLE connective is truth-functional.

But this is demonstrably not the case. Let "Phi" be "is metaphysically necessary". Let S be *1+1=2* and let S* be *Smith is bald*. TF shows *either* that *Smith is bald* is metaphysically necessary *or* that "necessarily" induces opacity. Obviously *Smith is bald* is not metaphysically necessary. So we must say that "Necessarily" induces opacity. Thus, "1+1=2" means one thing on its own, and a completely different thing in "necessarily 1+1=2", and yet a different thing in "it is necessarily the case that necessarily 1+1=2".

Some hearty spirits have maintained that "necessarily" induces opacity. But it is not a plausible position. And, I think, it is demonstrably false. Let "necessarily*" have the same meaning as "necessarily", except that, by our stipulation, any sentence S falling within the scope of "necessarily*" bears the proposition that it customarily (in truth-functional contexts) bears. If I say "necessarily* 1+1=2" I am saying, of the proposition customarily borne by "1+1=2", that it is necessary. Surely there is some proposition to exactly that effect.

An analogue of TF shows that either *Bob is bald* is necessary* *or* that "necessarily*" induces opacity. We know the latter is false. And we know, on independent grounds, that the former is false. So TF must be fallacious.

Here is another example of a connective that shows TF to be fallacious. The fact that grass is green is causally explained, at least in part, by the fact that grass contains chlorophyll. Let EXP be a two-place sentence-connective that we define as follows. "EXP (S, S*)" is true just in the

following holds. Let P and P* be the propositions *customarily* borne by S and S*. Suppose that P and P* are both true. Either P*, or one of the propositions composing it, provides at least a partial causal explanation of either P or one of the propositions composing it So

(e1) EXP (grass is green, grass contains chlorophyll)

is true. Plainly "EXP" is PSLE. For if we replace either constituent sentences with a logically equivalent sentence, what results is plainly true, given how we defined "EXP":

(e2) EXP (grass is green, 1+1=2 and grass contains chlorophyll),

is true. And so is

(e3) EXP (grass is green and 1+1=2, grass contains chlorophyll).

Given how we've defined it, "EXP" clearly does not induce opacity. After all, it is the essence of "EXP" that sentences falling in its scope bear their *customary* meanings.

At the same time, PSLE is plainly *not* truth-functional.

(e4) EXP (grass is green, Mozart was a great composer)

is obviously false. The fact that Mozart was a great composer does not, by any stretch, have any explanatory relevance to the fact that grass is green. And I'm pretty sure that everyone – even Quine, who puts forth TF (or, at least, a very similar argument) – would agree.

Conceivably, an advocate of TF might say that nothing is explanatorily relevant to anything, and that what TF establishes is therefore not absurd. This would obviously be desperate. Even if true, it wouldn't really speak to the larger issue. For we could obviously construct temporal or moral analogues of "EXP"; and the explanatory nihilism just stated wouldn't have any relevance to these.

The natural conclusion is to say that TF is fallacious. It is *not* true that PSLE connectives are always truth-functional.

The question is: what is the problem with TF? The problem, I

submit, is that it relies on SS.

Let us start with some uncontroversial points. (2) contains the definite descriptions: "the class of all things x such that (x=x and S)". (3) contains the definite description: "the class of all things x such that x=x and S*". Both contain the expressions "the class of all things x such that x=x".

TF assumes that "S" is logically equivalent with "(the class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x)." It also assumes that "S*" is logically equivalent with "(the class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x)."

Trivially, either definite descriptions are terms that refer to individuals or they aren't. If they aren't, they're Russellian quantifiers. Let's consider each case.

If they are Russellian quantifiers, then TF fails miserably. The transition from (2) to (3) no longer goes through. Let S be the sentence "Bach wrote fugues". Let S* be the sentence "Stalin was evil". In that case, (2) and (3) become:

(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and C=C*.

(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and C=C*.

(2_R) and (3_R) are definitely *not* logically equivalent. So the fact that "Phi" is PSLE does *not* warrant the transition from (2) to (3), at least not if definite descriptions are read as Russellian quantifiers.

One might say:

Obviously (2_R) and (3_R) are not logically equivalent. But this doesn't mean that the transition from (2) to (3) fails. It means that it fails only if "Phi" is given wide-scope, i.e. only if (2_R) and (3_R) in their entirety fall within the scope of "Phi". But if "Phi" is given narrow-scope, then the transition from (2_R) to (3_R) works. There are different narrow-scope readings. Perhaps

TF goes through on one of them.

Pick any narrow-scope reading you wish. The result will *in*validate one of the steps in TF. Here is the most obvious narrow scope reading of (2) and (3).

(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).

(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).

Let us put this reading into the context of the argument as a whole:

(1) Phi (Bach wrote fugues)
(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).
(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).
(4) Phi (Stalin was evil).

In that case, the steps from (1) to (2), and from (3) to (4), become completely invalid. For any objects C and C*, the proposition *Bach wrote fugues* is not equivalent to the proposition *C=C**. The same is true of *Stalin was evil.* So if we give "Phi" that particular narrow scope reading, then two much needed transitions simply fail; they are no longer guaranteed by the fact that "Phi" is PSLE.

Granted, there are other narrow-scope readings. But it is easy enough to verify that, on any of them, one of the steps in question becomes invalid. It would be a tedious and unnecessary exercise to go through each one of them.

It is clear what we must say. We must treat the definite descriptions

involved as expressions that refer to individuals. If we *don't* do this, then TF becomes a tissue of non-sequiturs.

But TF becomes a tissue of non-sequiturs if we *do* treat definite descriptions in that way. For the sake of argument, let the definite descriptions in TF be referring expressions. Let Omni be the class of all things x such that x=x. In that case, all of the definite descriptions refer to Omni.

Given this, remember what we established about definite descriptions. Let O be the unique thing having phi. *If* *the phi* is a singular term, *then* the proposition meant by *the phi has psi* is simply: *O has phi.* (And *if* the sense of *the phi* makes it into the proposition meant by that sentence, then *the phi* is a Russellian quantifier.)

As we've seen, Quine's argument requires that we treat definite descriptions as singular terms. Consequently, TF becomes:

(1) Phi (Bach wrote fugues).
(2) Phi (Omni=Omni).
(3) Phi (Omni=Omni).
(4) Phi (Stalin was evil).

In that case, TF is patently bankrupt. For no two of the following sentences are logically equivalent: "Bach wrote fugues", "Omni=Omni", "Stalin was evil". Consequently, the fact that "Phi" is PSLE doesn't guarantee *any* of the steps in TF.

There is only one way to save TF. We must say: two co-referring expressions can make *different* semantic contributions. So, for example, "the class of all objects x such that x=x" must make a *different* semantic contribution from "the class of all objects x such that (x=x and Bach wrote fugues)", even though *both* those expressions co-refer – even though they both refer to Omni. If TF is to work, we need (SS).

The other side of the coin is, once (SS) is granted, then TF does go through. But then we have the pernicious result that PSLE connectives are truth-functional.

Gödel's Slingshot

Recently a version of the Slingshot due to Kurt Gödel (Gödel 1946)

has received considerable attention.[13] Neale (1994, 2001) thinks that Gödel's Slingshot is, in some respects, superior to other versions of the Slingshot. This is a mistake.

Gödel's argument is meant to show that, given certain assumptions, it can be proven that all true sentences have the same "signification". It isn't clear whether, by "signification", Gödel meant "denotation" or "sense".

First of all, let *s(a)* be short for: the signification of *a*.

Here are the three assumptions Gödel's argument explicitly uses. (The asterices are quasi-quotes.)

G1: if *a=b* is true, then *s(a)*=*s(b)*.

G2: There is a function f such that for every a, b such that *a=b* is well-formed, *s(a=b)*=f(s(a), s(b)).

G3. s(*…a…*)=s(*the unique thing x such that x=a and…a…)*).

Gödel's argument involves a fourth premise. That promise is often not stated as a premise (it is usually just taken for granted):

G4: *Fa* is logically equivalent with *a=the unique thing x such that x=a and Fx.*

On the basis of G1-G4, Gödel an argument to the effect that any two true sentences have the same "signification". We don't need to look at that argument itself – let us concede that it goes through, given the premises. (See Neale 2001 for a careful reconstruction of the argument.) Let us now look at the premises.

First of all, we must clarify the third premise. Consider the sentence "Socrates is bald". G3 says that "Socrates is bald" and "the unique thing x such that x=Socrates, and x is bald" have the same "signification". As we've noted, there are two ways understanding the term "signification" here. It could mean either denotation or it could mean meaning. So G3 is ambiguous between the following two statements.

(m) "Socrates is bald" and "the unique thing x such that x=Socrates,

and x is bald" have the same *meaning*, i.e. they encode the same proposition.

(d) "Socrates is bald" and "the unique thing x such that x=Socrates, and x is bald" have the same *referent.*

Definite descriptions are either singular terms or quantifiers. It will turn out that both (m) and (d) are false regardless of how we treat definite descriptions.

First, suppose that definite descriptions are quantifiers. In that case, "the unique thing x such that x=Socrates, and x is bald" means: *something is uniquely identical with Socrates, and x is bald.* This proposition is not *identical* with the proposition *Socrates is bald.* They are logically equivalent – but not *identical.* The one proposition is an existence-claim; the other is not. The one proposition has the concepts of identity and uniqueness as constituents; the other does not. Those propositions have different compositions. Also, logical equivalence here would have to be *proved*. It would be an easy proof. But that doesn't matter. Where equivalence has to be proved, then (it seems to me) we are dealing with distinct propositions. So if definite descriptions are treated as quantifiers, then "the unique thing x such that x=Socrates, and x is bald" and "Socrates is bald" do *not* mean the same thing, and (m) is false.

Let us deal with (d). Once again, suppose that definite descriptions are treated as quantifiers. In that case, as we just saw, "the unique thing x such that x=Socrates, and x is bald" and "Socrates is bald" encode different propositions. In that case, (d) is false if sentences denote propositions. So for (d) to be true, it must be assumed that sentences do not denote propositions. But this cannot be assumed in this context. In the context of an argument purporting to establish what sentences denote, it cannot be *assumed* that they don't denote propositions.

To sum up, if definite descriptions are treated as quantifiers, then (m) becomes false and (d) becomes question-begging; so G3 becomes either false or question-begging, depending on whether we take "signification" to mean denotation or meaning. So if G3 is to be true, then definite descriptions must be treated as singular terms. So for the sake of argument, let us suppose that they are in fact singular terms.

In that case, G4 becomes false. We've already seen that *if* *the phi* is a singular term referring to some object O, then *the phi has psi* means: *O*

has psi. Given this, suppose that "the unique thing x such that x=Socrates" is a singular term. In that case, it refers to Socrates. So, for any property phi, *the unique thing x such that x=Socrates has phi* means: *Socrates has phi.* So, for example, "the unique thing x such that x=Socrates snored" means: *Socrates snored.* And "the unique thing x such that x=Socrates has phi=Socrates" means: *Socrates=Socrates.*

Let me be as clear as possible. We've seen that *if* *the phi* is a singular term referring to O, then *the phi has psi* means: *O has phi.* So *if* "the unique x such that x=Socrates" refers to Socrates, then when that expression occurs in a sentence, it is Socrates, and Socrates alone, that makes it into the corresponding proposition. So "the unique x such that x=Socrates snores" means simply *Socrates snores,* and "Socrates is identical with the unique x such that x=Socrates snores" means simply *Socrates=Socrates.* In that case, "Socrates was bald" and "the unique thing x such that x=Socrates, and x was bald" are *not* logically equivalent. One means *Socrates was bald.* The other means *Socrates=Socrates.*

Let us generalize this point. If definite descriptions are singular terms, then for any property phi, *a is identical with the unique x such that x has phi* means simply: *a=a.* So *a is identical with the unique x such that x=a and Fx* means simply *a=a.* Of course, what is meant by *Fa* is *not* typically logically equivalent with *a=a.* (They would be equivalent only for quite specific values of F and a.) So *if* definite descriptions are singular terms, then *Fa* is *not* equivalent with *a is identical with the unique x such that x has phi*. Let F be the property of baldness, and let a be Socrates. The one means *Socrates=Socrates.* The other means *Socrates is bald.* So *if* definite descriptions are singular terms, then G4 is false.

Some other Slingshots

On the basis of Slingshot-style, some philosophers actually *have* accepted that there are no facts (or only one fact) and no propositions, and that our intuitions about modality are spurious. There were independent motivations (not to say justifications) for these views. Empiricists have always held meaning (propositionality) and modality in contempt. (Interestingly, these notions are interconnected.) And the notion of a *fact* is famously problematic. Under what circumstances are x and y different facts? Let H be some heart. Consider some instance of H's beating. That beating

involves innumerable sub-atomic events. Let $E_1 \ldots E_n$ be those events. Is the occurrence of $E_1 \ldots E_n$ identical with the beating of that heart? Is the fact that triangles have three sides identical with the fact that a thing has three sides iff it a closed, straight-edged figure such that any two, but not all three, of its sides intersect?

Do facts decompose in the same way as the propositions that express them? Consider the proposition *H is beating at time t*. That proposition has a fairly simple decomposition. But the event of H's beating is obviously far less simple. That suggests that propositions and facts are altogether different sorts of things. But how are we to individuate facts except in terms of the propositions that affirm them?

These questions are not easy to answer. Some say that there is no principled way of answering them and that, consequently, the concept of a "fact" is an incoherent one. So we should do away with facts. The Slingshot, evidently, gives this antipathy a logical, as opposed to purely emotional, basis.

There is another dimension to this. A number of philosophers and psychologists think that the notion of a *representation* is bankrupt. The later Wittgenstein is one of the more vigorous spokespersons for this viewpoint. Davidson added a new wrinkle to this debate. He argues that the notion of a *representation* is closely bound up with the notion of a fact. In his view, if there are no facts, or only one fact, then there are no representations. If the fact that snow is white collapses into the fact that grass is green, then the corresponding representations also collapse, which would make representations useless.

And, of course, there is the idea that sentences refer to truth-values. Intuitively, this idea seems false. The relation that holds between "snow is white" and the truth-value True doesn't seem at all like the relationship that holds between "the truth-value True" and the truth-value True.

But as Frege observed, there are good reasons to think that sentences *do* refer to truth-values. That notion seems to be an inevitable consequence of an indubitable semantic principle – the principle of compositionality. (Incidentally, I agree with Frege that compositionality must be respected at all costs. In fact, we will see that, given a correct conception of reference, compositionality is truistically true; and denying it would be like denying that triangles have three sides.) Once again, the Slingshot comes to the rescue: it supports an idea that seems false but that we want to be true, because it

accommodates a powerful body of theory.

This is all well and good. There are two problems. First, the Slingshot is spurious. (We've seen one reason why – we are about to see others.) Second, if we grant the (demonstrably false) assumptions needed for the slingshot to go through, we can "prove" things that *nobody* would accept, no matter their theoretical commitments. Let us now show this.

Obviously different physical objects are involved in different events. One set of objects was involved in Kennedy's assassination (a certain bullet, a certain car, and so on). Another was involved in Mozart's composition of the Jupiter Symphony (a certain quill pen, sheets of paper, and so on).

When I say "involved", I mean directly involved; I am not talking about remote causal involvement. So the bullet that killed Kennedy was directly involved; but the chopping down of the tree that yielded the wood composing Oswald's rifle is not directly involved.

Of course, analytic sentences do not describe events at all. So "Kennedy was assassinated" and "Kennedy was assassinated and 1+1=2" describe the same event. The second encodes information not encoded in the first. But that information has nothing to do with *events*; it has to do with timeless relations among numbers.

The core of the argument is given by the following sequence of expressions:

(a) The set of physical objects involved in the event described by the following sentence: Kennedy was assassinated.

(b) The set of physical objects involved in the event described by the following sentence: The class of all objects x such that (x is identical with x and Kennedy was assassinated) is identical with the class of all objects x such that (x is identical with x)

(c) The set of physical objects involved in the event described by the following sentence: The class of all objects x such that (x is identical with x and Mozart composed the Jupiter symphony) is identical with the class of all objects x such that (x is identical with x)

(d) The set of physical objects involved in the event described by the following sentence: Mozart composed the Jupiter symphony.

(a) and (b) co-refer by LL and CR. (b) and (c) co-refer by CR and LL. (c) and (d) co-refer by LL and CR

Of course, if (a) and (b) are to be logically equivalent, (SS) must be true. For if (SS) is false, then (b) and (c) both become: *Omni is Omni*, which is not logically equivalent to *Kennedy was assassinated* or to *Mozart composed the Jupiter Symphony*. For similar reasons, (c) and (d) are equivalent only if SS holds.

So if we accept the assumptions that drive the Slingshot – (LL), (SS), (CR) – we must accept that (a) and (d) co-refer. The set of physical objects directly involved in Kennedy's assassination is identical with the set involved in Mozart's composing of the Jupiter symphony. The set of physical objects directly involved in *any* event is identical with the set of physical objects directly involved in any other.

This is obviously false. Also, it implies that there is only one physical object. Presumably x and y are *different* physical objects iff they are involved in different events. So there is only one physical object. This is what the Slingshot proves.

Here is a similar argument. Some philosophers distinguish between constitution and identity. The statue is not *identical* with the clay because you can destroy the statue without destroying the clay. But the clay obviously *constitutes* the statue.

Everyone agrees that every physical entity is *either* identical with, *or* constituted by, displacements of mass-energy. Your car is *constituted* by a long, complicated series of mass-energy displacements. Let S be a sentence that describes exactly that series. Your favorite shirt is constituted by some entirely distinct series of mass-energy displacements. Let S* be a sentence that describes exactly that series.

For obvious reasons, the series of mass-energy displacements that makes S true is identical with the series that makes S, conjoined with "triangles have three sides", be true.

(A) The set comprising all and only the mass-energy displacements that make true the sentence: S.

(B) The set comprising all and only the mass-energy displacements that make true the sentence: The class of all objects x such that (x is identical with x and S) is identical with the class of all objects x such that (x is identical with x)

(C) The set comprising all and only the mass-energy displacements that make true the sentence: The class of all objects x

such that (x is identical with x and S*) is identical with the class of all objects x such that (x is identical with x)

(D) The set comprising all and only the mass-energy displacements that make true the sentence: S*.

The set comprising all and only the mass-energy displacements that make up your favorite shirt has exactly the same members as the set of such displacements that make up your car. Any two objects are constituted by exactly the same mass-energy displacements. Given any two physical objects, they are either identical, or one of them constitutes the other.

Obviously, the Slingshot proves *too* much. There must be something wrong with it. We've seen one thing that is wrong with it – it presupposes an incoherent conception of reference. But other things are wrong with it. In fact, it can be shown that *if we accept* the conception of reference that the Slingshot needs, that actually *makes* the Slingshot fail. The Slingshot fails on its own terms.

Most Slingshots involve (LL) and (CR). There are a number of problems here. (CR) is, I believe, correct. In fact, I think it holds *without restriction.* The problem is that the semantics that is needed to make (CR) be true also undermines the Slingshot. As we will see, the semantics that the Slingshot needs require that (CR) be true only in certain contexts. So, somewhat ironically, the Slingshot needs (CR) to be less than strictly true; it needs a certain *weakening* of (CR) to be true. But we will see, first, that the Slingshot cannot non-question-beggingly use *any* version of (CR) and, second, that there is no principled weakening. (CR) is like butterfly wings: if you touch it, it is destroyed. When we try to abridge (CR), it becomes worthless, and it drags down any semantic superstructure that it is meant to support. I will argue this in a moment.

The other problem concerns (LL). There are several problems here. First, (LL) cannot be non-question-beggingly used in the needed contexts. Second, any arguments that might be given for (LL) turn out to be fallacious; and the intuitive motivations for it dissolve.

When we attend to some key innovations of contemporary semantics, e.g. the distinction between type-semantics and also token-semantics and also that between sentence-logic and proposition-logic. (Any token of the *sentence* "I exist" must be true; so, from one viewpoint, any such *sentence*-token is "logically equivalent" with a token of "1+1=2". But the

proposition encoded in a token of the former is contingent.)

§ Right now let us talk about (LL), beginning with a discussion of why the Slingshot, in any of its embodiments, cannot use it.

(LL) Logically equivalent sentences co-refer.
(CR) Intersubstitutions of co-referring terms preserve reference.

Given this, consider the following sequence of sentences:

(1) "Mozart wrote music."
(2) "The class C such that x is a member of C iff (Mozart wrote music and x=x) is identical with the class C such that x is a member of C iff (x=x)."
(3) "The class C such that x is a member of C iff (Smith is tall and x=x) is identical with the class C such that x is a member of C iff (x=x)."
(4) "Smith is tall".

(1) and (2) are supposed to co-refer by (LL). (3) and (4) are supposed to co-refer by that same principle.

(2) and (3) are supposed to co-refer by (CR). In this case, the co-referring terms in question are: "the class C such that x is a member of C iff (Mozart wrote music and x=x)" and "The class C such that x is a member of C iff (Smith is tall and x=x").

So (1) and (2) are supposed to be logically equivalent; so are (3) and (4); and these supposed equivalences are crucial to the Slingshot.

We know that the transformations which lead from (1) to (4) preserve *reference*.

The proposition expressed has clearly *not* been preserved -- (1) and (4) obviously express different propositions. So sentences don't refer to propositions. The one semantically significant thing that (1) and (4) have in common is *truth-value.* So it is natural to identify a sentence's referent with its truth-value.

When one is trying to prove that sentences refer to truth-values, one

is also, at least by implication, trying to prove two other things. First, sentences do refer to *something* or other. Second, they *don't* refer to propositions. What does "snow is white" refer to? One possible (not to say correct) answer is: the proposition that snow is white. Later I hope to show that there is some demonstrative support for this view. If you say "if snow is white, then snow is not green", you are asserting that a certain relation – a relation of consequence -- holds between two propositions (*that snow is white* and *that snow is not green*). It is thus natural to see the "if" (or "if…then…") as something like a two-place predicate that ascribes the consequence-relation to pairs of propositions, and to see sentences falling within the scope of "if" as denoting those propositions.

In any case, someone who is putting forth an argument that sentences refer to truth-values certainly cannot *take it for granted* that they don't refer to propositions. Such an argument would be question-begging. Now there are perhaps good grounds for thinking that sentences *don't* refer to propositions. But those grounds would have to be provided independently. And there is no guarantee that those grounds wouldn't *also* work against the idea that sentences refer to truth-values. (In fact, that is exactly what they do, as we will see.) So the use of (LL) in the context of the Slingshot is simply question-begging.

§ Let us now talk about (CR). We will focus on the same version of the Slingshot as before.

Supposing (as we are) that SS is true, then CR does not hold universally. As many have observed, it apparently fails in so-called "non-extensional" contexts.[14] Suppose John knows that Franklin invented bifocals and also that Franklin snored; but suppose John doesn't even know what a post-master general is.

(i) "John thinks that the inventor of bifocals snored".

(ii) "John thinks that the first post-master general snored".

The first is true, the second is false. (ii) seems to be what results when a referring term is replaced with a co-referring term.[15] Now the reason (i) and (ii) differ in truth-value appears to be that the underlined clauses refer to *different* propositions. So it appears that, in this case, replacing a referring

term with a co-referring term resulted in a change of reference.

For this reason, it is widely held that CR fails in so-called "non-extensional contexts". So, strictly speaking, what is true is not CR, but rather:

(CR1) In *extensional* contexts, intersubstituting co-referring terms doesn't change reference.

Obviously the Slingshot cannot use CR as a premise, since it is false; what it uses is presumably CR1.

But the question now is: what does "extensional" mean? A standard way of defining that term is this:

(EX1) A context is extensional iff, in that context, intersubstituting co-referring terms doesn't change truth-value.[16]

But if (EX1) is the right definition of "extensional", then CR1 is false, and thus cannot be used by any argument. Consider:

(a) "It is true that the inventor of bifocals snored".
(b) "It is true that the first post-master general snored".

In each of (a) and (b), the definite description occurs in a context where intersubstituting co-referring terms doesn't change truth-value. Each underlined clause is an expression that refers to a proposition Those clauses refer to different propositions. But the one clause is (on the face of it[17]) what results when a referring term in the other is replaced with a co-referring term. So if (EX1) is the right definition of "extensional", then CR fails in *extensional* contexts – so CR1 becomes false. Of course, if CR1 is false, then it is of no use to the Slingshot (or any other argument).

Supposing (as we still are) that SS holds, here is the principle that underlies any *correct* abridgment of (CR). Let E be an expression that *rigidly* denotes a meaning (for example, a proposition or concept); and let e be some referring term that occurs as a part of E. In that case, replacing e with a co-

referring term e* *can* change the referent of E.

The expression:

(A) "that the inventor of bifocals snored"

refers to a proposition. That seems to be why, when we replace the definite description with a co-referring definite description, the result is an expression that refers to a *different* proposition:

(B) "that the first post-master general snored".

Or consider: "the concept of being identical with the inventor of bifocals". When we replace the definite description with a co-referring definite description, we produce an expression that refers to a different concept: "the concept of being identical with the first post-master general". So, at first glance, it looks as though, in the context of an expression E that refers to a proposition or concept, intersubstituting co-referring terms can change the referent of E.

But this is not *quite* accurate. Some expressions referring to propositions or concepts tolerate such substitutions:

(C) "the favorite proposition the author of Waverly".

If we replace "the author of Waverly" with a co-referring definite description, what results is an expression that co-refers with (C):

(D)"the favorite proposition of Sir Walter Scott".

Notice that (C) and (D) are *non-rigid* designators, whereas (A) and (B) are rigid designators. Here is the general principle that appears to fall out of these considerations:

(S) If E is an expression that *rigidly* designates a proposition or concept, and e is a referring term that occurs as a proper part of E, then replacing e with a co-referring term e* may result in an expression that has a different referent from E.

For example: (A) rigidly refers to one proposition; when we replace "the inventor of bifocals" with a co-referring definite description, what results is (B), which rigidly refers to some other proposition.

Now we can close the argument. If (S) is true, then (CR) is *unavailable* to someone who is using the Slingshot to show that sentences refer to truth-values. If sentences *do* rigidly denote propositions, then replacing a referring term in a sentence S with a co-referring term *can* result in a sentence S* that does *not* co-refer with S. If "the inventor of bifocals snores" rigidly refers to the proposition it means, then replacing the definite description in that sentence with "the first post-master general" will result in a sentence that does *not* co-refer with the original sentence. So *if* sentences rigidly denote propositions, then (CR) is useless for the proponent of the Slingshot. So such a person must *assume* that sentences do not rigidly denote propositions. But this is not something that can be assumed in this context.

Further, as we discussed earlier, there is good reason to think that it *does* denote that proposition. Consider the sentence "if Mozart wrote music, then somebody wrote music ". What is going on here? This sentence affirms a relation between two positions (*that Mozart wrote music* and *that somebody wrote music*). It says that the one entails the other. Obviously it is the component sentences that are picking out those propositions, and it is the "if…then.." that is picking out the relation in question. Given that the component sentences are "picking out" the right propositions, it's a little hard to see how they could not be *referring* to it. Referring just *is* picking out – is it not? In any case, if (S) is true, then use of (CR) in the context of the Slingshot is question-begging.

A Point about Frege's Original Argument

Frege writes[18]:

"If we substitute a word in [a sentence] by another word with a different sense but the same nominatum [denotation], then this substitution cannot affect the nominatum of the sentence."

This sounds like a version of (CR): intersubstitutions of co-referring

terms preserve reference. But we've already seen the trouble with this assumption. Let e be a referring term that is a proper part of some other referring term. Suppose we replace e with a co-referring term e*, and let E* be the resulting host-expression. If E is an expression that rigidly denotes a proposition or concept, then E and E* may not co-refer. So if intersubstitutions of co-referring terms are to preserve *sentence*-reference, it must be assumed that sentences are not expressions that rigidly denote propositions. But this is not something that can be assumed in such a context. So to the extent that it relies on that assumption, Frege's original argument is vitiated.

§ Here there is a very good point to be made *against* the argument just given:

It may not be easy to say in virtue of what, precisely, a context is non-extensional. But we know that some contexts are extensional and that some are not. We also know two other things. First, (CR) holds in extensional contexts. Second, the substitutions that occur in the Slingshot occur in extensional contexts. So everything you've said so far is irrelevant. (CR) does preserve reference in precisely the contexts where the Slingshot needs it to do so.

This brings us to a much darker problem connected with (CR): a problem the resolution of which is crucial to the development of any adequate semantics.

Let me outline the argument I'm about to give. As we've noted, the Slingshot needs it to be possible to change meaning by inter-substituting co-referring terms. In other words, the Slingshot needs (SS). At the same time, if (SS) is granted, then it immediately follows that some contexts are *not* extensional, and this invalidates some of the steps in the Slingshot. For the sake of argument, let's suppose that there are non-extensional contexts. There are a few different ways of defining the term "non-extensional." But on any of those definitions, (CR) is voided, and all that is left is the innocuous triviality that intersubstituting co-referring terms preserves reference except

when it changes reference. This is not acceptable. (CR) *must* be correct; for (CR) is equivalent to the principle of compositionality, and compositionality is surely correct.[19]

Bottom line: if we grant that there are non-extensional contexts, then we must throw out CR and, therefore, the principle of compositionality. Since (CR) is correct, it follows that there are no extensional contexts. And if there are no extensional contexts, then it follows that (SS) is false. So (SS) entails that compositionality is false, and is therefore itself false.

We will find that the data which motivate the view that there are such contexts can easily be accommodated within a purely extensionalist semantics: the trick is to look very carefully at what is done *pre*-semantically. Consider the following pair of sentences:

(a) "John believes that four is greater than three":
(b) "John believes that the number of stomachs had by a normal cow is greater than three".

There is no doubt that the proposition *communicated* by (a) is very different from that communicated by (b). The proposition *communicated* by (b) is something like: *John believes that there is some number n such that n is the number of stomachs had by a normal cow, and n is greater than three.* The proposition *communicated* by (a) is simply: *John believes that four is greater than three.*

Everybody agrees that what is *communicated* by tokens of (a) will differ, and can therefore differ in truth-value, from what is communicated by tokens of (b). The question is: is this difference in respect of what is communicated a matter of *literal meaning*? Is the *literal* meaning of (a) different from that of (b)? Somebody who believes that there are intensional contexts says "yes". Let us refer to that doctrine as "intensionalism".

My own answer to that question is this. First, we must distinguish types from tokens. The semantics of the *type* associated with (a) is very different from the semantics of the type associated with (b). But I believe that *tokens* of (a) have the same literal meanings as *tokens* of (b). I believe that the massive differences in respect of what is communicated derive, not from semantics, but from *pre*-semantics.

But right now we are not going to deal with that issue: we are going to confine ourselves to demonstrating one thing: (SS) entails that

compositionality is false, and is therefore itself false.

For the sake of argument, let us assume that *intensionalism* is correct.

In this context, we are going to have to ignore the distinction between types and tokens. It becomes hard to maintain an intensionalist viewpoint if one remains too conscious of that distinction.

A good starting point is Kenneth Taylor's (1998: 181) excellent discussion. He says that the "hallmarks" of intensionality are the following:

(1) 'Intensional sentence connectives and operators are not *truth-functional*."

The occurrence of "the inventor of bifocals snored" in "John believes that the inventor of bifocals snored" is not truth-functional. If you replace that occurrence with another true sentence, the result may be false.

(2) "Intensional contexts exhibit *a tolerance for emptiness.*"

"John believes that the fountain of youth is in Florida" has a truth-value, even though "the fountain of youth" lacks a referent.

(3) "Intensional contexts block *the free substitution of co-extensive expressions.*"

See (a) and (b) above.

(4) "Intensional contexts *block the free exportation of quantifiers* and apparently *restrict the inner reach of external quantifiers.*[20]

"John believes that the fountain of youth is in Florida" is true. But "there is some x such that x is the fountain of youth, and John believes that x is in Florida" is false.[21]

Each of (1)-(4) gives us a *symptom* of something. What is that underlying something? What is the essence of intensionality? The answer is simple: intensionality is non-truth-functionality. The list above collapses into (1). Let s be any sentence that occurs in some other sentence S. If s occurs

non-truth-functionally, s will have some or all of the symptoms described. If s occurs truth-functionally, it will have none of the symptoms described.

Consider:

(d) "John believes that the fountain of youth is in Florida."

The underlined sentence occurs non-truth-functionally. Emptiness is tolerated. ("The fountain of youth" is empty, but the sentence has a truth-value.) Exportation is blocked: you cannot infer that, for some x, John believes that x is in Florida. Intersubstituting co-extensive terms isn't possible: if you replace "the fountain of youth" with "the sole member of the class of objects that are not self-identical", what results is false.

Now consider the sentence:

(e) "It is true that four is greater than three."

The occurrence of "four is greater than three" doesn't have any of the three symptoms. Intersubstituting co-extensive expressions is allowed: you can replace "four" with any co-referring expression of your choice, and the result will be true. Exportation is allowed: You can infer *for some n, it is true that n is greater than three.* Emptiness is not tolerated: replace each, or both, of "four" and "three" with an empty expression, and what results is truth-valueless (or false). (Consider "it is true that the highest prime is the highest prime".) The underlined clause occurs truth-functionally: replace it with any other expression denoting a true proposition, and the result is true.

Another example might be in order. Consider the sentence:

(f) "either Smith is a composer or grass is green".

Each underlined sentence is an extensional context, and each occurs truth-functionally.

So extensionality is truth-functionality. Any analysis not identical or equivalent with this one is false. Consider the following, apparently reasonable view:

(*) Let s be a sentence that occurs as a proper part of some other sentence S. If s *denotes a proposition,* then s is non-extensional.

The idea would be this. Consider (d) again. The underlined clause denotes a proposition, and it is a non-extensional context. Hence (*).

But we've already seen the problem with (*). Consider (f). Each of the underlined sentences obviously picks out a proposition. So it's a little hard to see why they don't simply *refer* to propositions. If "Smith" is a part of a sentence S, then the proposition meant by S *ipso facto* has Smith as a component. For this reason, "Smith" refers to Smith; indeed, for this reason, "Smith" refers to Smith in the strongest possible sense. If "grass is green" occurs as a part of a sentence S, then the proposition *that grass is green* is ipso facto a component of the proposition meant by S. So the relation between such an occurrence of "grass is green" and the proposition *that grass is green* is identical with the relation between "Smith" and Smith. So, presumably, occurrences of that sentence refer to that proposition in the strongest possible sense. So (*) would seem to be quite false. In any case, a very counter-intuitive analysis of reference would have to be produced to validate (*). So far as I can tell, (*) is the only live alternative to our analysis of extensionality. Thus, we may accept that analysis: extensionality is truth-functionality.

But here is where the problems begin. It is clear that in:

(e) "It is true <u>that four is greater than three</u>,"

the underlined clause is doing the exact same thing that it is doing in:

(a) "John believes <u>that four is greater than three</u>".

In each case, the underlined clause is picking out the proposition *that four is greater than three.* We must conclude that, in each case, the occurrence of "four is greater than three" is doing the very same thing.

In (a), the occurrence of "four is greater than three" has all the symptoms Taylor mentions. In (e), the corresponding occurrence has none of those symptoms. Since the occurrence in the one is doing precisely the same thing as the corresponding occurrence in the other, we cannot say the presence of these symptoms in (a) has anything to do with some kind of *semantic* shift. The presence of those symptoms must derive from the material *surrounding* that occurrence. The occurrence *itself* is disease-free.

Now we are in a position to see why intensionalism is inconsistent with the principle of compositionality. As we just saw, there is no denying that the occurrence of "four is greater than three" in (a) is semantically just like its counterpart in (e). The intensionalist must admit this. If this intensionalist is right, then in (e), we must *not* say that the occurrences of "four" and "three" are referring to numbers. If the intensionalist is right, then replacing the occurrence of "four" in (e) with "the number of stomachs had by the average cow" results in a false sentence. So if the intensionalist is right, then for reasons discussed earlier, it would be inconsistent with Leibniz's Law to hold that, in (e) those occurrences refer to numbers. In general, if the intensionalist is right, then occurrences of nouns (and other referring terms) in *any* epistemic or modal context will fail all the tests that their counter-parts in truth-functional contexts will pass: they will fail the exportation-test, the substitution-test, and the must-have-a-referent test.

Since the occurrences of "four" and "three" in (a) are doing exactly the same thing they are doing in (e), we must say that, in (a), those occurrences don't refer to numbers. Remember what we established. The occurrence of "four is greater than three" is doing the very same thing in both (a) and (e). If the intensionalist is right, then in (a) those occurrences don't refer to numbers. So if the intensionalist is right, then they don't refer to numbers in (e). By itself, this seems very strange: surely the occurrence of "four" in "it is true that four is greater than three" refers to the number four.

This point can be taken further. If the intensionalist is right, then the occurrences of "four" and "three" in "that four is greater than three" don't refer to numbers. For reasons we've seen, the intensionalist must say that, in (a), the corresponding occurrences don't refer to numbers. We know that in "that four is greater than three", those occurrences are doing exactly the same thing which they are doing in (a) – after all, (a) is built out of that expression. So the intensionalist must say that those occurrences don't refer to numbers in "that four is greater than three".

This is hard to believe.

For exactly similar reasons, the intensionalist must say that in "that John snores", there is no expression referring to John.

Again, this is hard to believe.

This can be taken further. Suppose we accept compositionality (as I do). In that case, we must say that the referent of "that the number of stomachs in the average cow is greater than the number of sides of a triangle"

is a function of the referents of its components. For reasons we've considered, if one accepts intensionalism, one can*not* say that the occurrences of the definite descriptions in that expression denote numbers. After all, an intensionalist will say that "that the number of stomachs in the average cow is greater than three" denotes a different proposition from "that four is greater than three." That is his explanation of why the proposition conveyed by (a) is true whereas the proposition conveyed by "(b) is false. The difference in *sense* between "four" and "the number of stomachs had by the average cow" So if one accepts intensionalism, then one must say that the referent of that expression is a function of the *senses* of the referring terms occurring in it. So an intensionalist must say that, in that expression, the constituent nouns denote *senses*, not numbers. In general, if one is an intensionalist, one must say, for any expression P, the expressions occurring in *that P* denote *senses*.

This means that, for any object x that is *not* a sense, there is no expression of the form *that P* comprising an expression denoting x.[22] Suppose you want to produce an expression of the form *that P" contains an expression referring to John. One would think this easy: "that John snores", "that John is tall", and so on. But the intensionalist says: "No – in each case, what occurs is an expression denoting a *sense* that picks out John; there is no expression denoting John".

So if the intensionalist is right, there is no expression (in natural language) denoting a proposition that contains an expression denoting a rock, a person, or a number.[23]

This is hard to believe.[24]

Let us move on. We know that in:

(g) "four is greater than three"

the occurrences of "four" and "three" refer to numbers. And we know the same to be true of the corresponding occurrences in:

(h) "either four is greater than three or Mozart was not a composer".

If intensionalism is correct, then it must be able to answer the question: "Why do the occurrences of "four" and "three" in (h) refer to numbers, while the corresponding occurrences in (a) and (e) do not thus

refer? What is the principle involved?" There is no way to answer that question. The only answer to be given is this: "In (a) and (e), those occurrences occur within the an expression that denotes a proposition, whereas this is not the case with the corresponding occurrences in (h)." But this answer is false, as we've seen.

An intensionalist must insist that the occurrences of "four" and "three" *do not* refer to the same things as their counterparts in (a) and (e): and *on the basis of this very fact,* he might deny that the component sentences in (h) refer to propositions. But this is obviously just an epicycle. We know that the component sentences in (h) *pick* out propositions. This is a hard datum. Nobody would deny it (though some would deny that this is enough for *bona fide* reference). According to the intensionalist, the occurrences of "four" in (a) and (e) don't refer to a number. This is not a *datum*: it is a requirement of a theory. Further, that requirement is *itself* opposed to some sturdy pre-theoretic intuitions: it is *itself* an epicycle. So the intensionalist is denying hard data to keep a theory afloat: he is adding epicycle to epicycle. This is a classic symptom of what Imre Lakatos referred to as "degenerating research program".

But even if one denies that the component sentences in (h) refer to propositions, it is very hard to believe that, in (e) and in "that three is greater than four", the occurrences of "four" and "three" don't refer to numbers. As we've seen, the intensionalist *must* deny that they so refer. That by itself is a major strike against intensionalism. Later we will consider a clever argument of Davidson's that supports this intuition.

Here is another way of looking at it (I am anticipating some points to be made later about sentences). Consider (g). The proposition expressed by that sentence is *that four is greater than three.* But (g) is different from the expression:

(i) "That four is greater than three".

(g) *affirms* the proposition in question; (i) does not. The expression "that it is true that four is greater than three" is no more an affirmation than is (i). Denoting never amounts to affirming. To turn (i) into an affirmation, we need to give it assertoric *force.* In English, there are different ways to do this. We can drop the "that". Or we can prefix it with an "it is true". Suppose we drop the "that". We then end up with (g). From a purely orthographic

standpoint, (g) is simpler than (i). But semantically (g) is more complex. (i) simply picks out a proposition; it doesn't do anything with it. But (g) not only indicates a proposition: it affirms it. So (g) involves picking out that proposition *and* affirming it. So (g) does everything that does (i) *and then some.* So (g) is semantically *more* complex than (i), even though phonetically it is less complex. Consider a language English* that is just like English except for one thing: in English*, phonetic surface structure never hides semantic structure the way it hides it in English. The English*-translation of (i) would comprise an occurrence of something that indicated the proposition *that four is greater than three.* It would comprise an occurrence of an expression like "that four is greater than three". What I am suggesting is that the *non*-occurrence of "that four is greater than three" in (g) is an orthographic epiphenomenon – it is a simple case of ellipsis. Language is replete with ellipsis – with "aphonic" or "phonetically unrealized" components. Here is a trivial example. Somebody asks you "who ate my cookie"? You say "Bob". What you are saying is elliptical for "Bob ate your cookie". Here is a less trivial example. Depth-grammarians have made a powerful case that in "John wants to go to the store", the occurrence of "to go to the store" is really elliptical for "that John go to the store". So semantically, though not phonetically, there are *two* occurrences of "John" in that sentence. For similar reasons, I believe that, in (g), there is an aphonic occurrence of "that four is greater than three".

There is some reason to believe that (g) is an *ascription* of truth to the proposition *that four is greater than three.* After all, (g) is an affirmation of that proposition; and it is hard to see how affirming a proposition could be different from ascribing truth to it. (i) does not itself ascribe truth to that proposition: it *denotes* a proposition that ascribes truth to it, but does not itself do that. If this is right, then it is reasonable to see (g) as comprising some constituent that picks out, and thus denotes, that proposition. How could (g) ascribe truth to that proposition if, semantically, it didn't have such a constituent? If an expression picks out x, then *ipso facto* that expression picks x out, or at least has a component that does. In any case, if this is right, then (g) *does* comprise an (aphonic) occurrence of "that four is greater than three" or, at any rate, some equivalent expression. And if *that*, in its turn, is right, then *whenever* an expression occurs in the context of a sentence, it occurs within the context of an expression that rigidly denotes a proposition. For reasons we've seen, this would mean that, if intensionalism is right, *no*

occurrence of "three" denotes a number: *nothing ever denotes anything other than a sense*. But that is absurd. So intensionalism is wrong.

One *could* say that I have the facts exactly the wrong way around. An occurrence of "four is greater than three" naturally has assertoric force. Adding a "that" to it *cancels* that force. So "that four is greater than three" *is* semantically more complex than "four is greater than three". The former comprises the latter, both semantically and phonetically, *and* it also contains something that neutralizes one of the components of the latter.

This is certainly a way to go; and we will consider it in due course. The problem is that there is no independent reason to believe that "that four is greater than three" comprises any kind of force-operator. And there are a lot of reasons to think it does not. Consider the sentence "is it true that four is greater than three?" This has interrogative force. If the proposal being considered is right, then that sentence comprises a component which has assertoric force. But there are a lot of reasons, known to philosophers and linguists alike, that sentences cannot have *two* forces. So we may conclude tentatively, but with some confidence, that the proposal in question is not correct. And, for the reasons given a moment ago, there is some evidence that, semantically if not phonetically, "four is greater than three" *does* comprise an occurrence of "that four is greater than three".

If this is right, then the occurrences of "four" and "three" *must* co-refer with their counterparts in (i) and (a) and (e). But we want to believe this, even if that analysis of (g) isn't right.

Let us sum up. For many reasons, it is quite clear that the occurrences of "four" and "three" in (e) denote the same things as their counter-parts in "four is greater than three". The intensionalist must say that in (e), they don't denote numbers – they denote *senses.* So he must say that in "four is greater than three" they don't denote numbers. But this is obviously false. So intensionalism is false.

Also, as we promised to show, intensionalism is inconsistent with compositionality. As we've seen, to hold onto compositionality, the intensionalist must say that, in (a) and (e), the occurrences of "four" and "three" don't refer to numbers. So, on the intensionalist view, compositionality holds *only if* those occurrences don't refer to numbers. Those occurrences obviously co-refer with their counterparts in "four is greater than three". The occurrences in the latter do refer to numbers. So their counterparts in (a) and (e) also so refer. So, if intensionalism is right, then the

condition needed for compositionality to hold is not met. Intensionalism is inconsistent with compositionality.

There is one last point. Consider:

(e) "it is true that four is greater than three".

The occurrence of "four is greater than three" has none of the four hallmarks of intensionality identified by Taylor. But, as we've seen, nobody – not even an intensionalist – would deny that, in (e), that occurrence has the same semantics as its counterpart in:

(a) "John believes that four is greater than three".

To save compositionality, the intensionalist must deny that, in (a), those occurrences refer to numbers. So he must deny that they so refer in (e). Thus, he must say that, semantically, the occurrence of "three is greater than four" in (e) is *intensional*. But the problem is: that occurrence has *none* of Taylor's hallmarks of intensionality. That occurrence passes all the tests passed by its counterpart in "four is greater than three" and failed by its counterpart in (a). So the intensionalist says that a context can be intensional even though it doesn't bear any resemblance at all to our paradigms of intensionality. So a context can be intensional even if it has none of the properties definitive of our concept of intensionality. This, it seems to me, borders on incoherence.

§ There is one other point to consider. We can't define an intensional context as one where (CR) doesn't hold, i.e. as one that doesn't tolerate intersubstitutions of co-referring terms. That would trivialize the principle of compositionality, as I would now like to argue. The principle of compositionality is equivalent with (CR). So if we defined an intensional context in the way proposed, then we'd have to say this: compositionality holds except in intensional contexts, and an intensional context is one where compositionality doesn't hold. In other words, compositionality holds – except when it doesn't. And (CR) holds – except when it doesn't. So compositionality/CR would be nullified if we defined an intensional context as one where (CR) doesn't hold.

We must therefore regard the exceptions to (CR) as *symptoms* of an

underlying conditional. What is that condition? The only viable answer is: (CR) fails to hold when the context in question (rigidly) denotes a meaning (a proposition or concept) of some kind. But then (CR) *never* holds since, relative to any but a very implausible conception of reference, any sentence can be seen as constituting such a context.

§ Intensionalists will say that the occurrence of "four is greater than three" in "it is true that four is greater than three" is intensional, even though it is truth-functional. But that position, we saw, is a non-starter. Given that fact, and given what we saw a moment ago, we may conclude that a context is extensional exactly if it is truth-functional.

But I feel that whether a sentence s occurs truth-functionally or not is a *symptom* of something deeper. In (e), *truth* is what is being ascribed to the proposition denoted by the underlined clause. In (a), the property of being believed by John is the property ascribed.

This might tempt us to say the following. Let s be a sentence that occurs in some other sentence S. s is an extensional context if S is an ascription of truth or falsity to s. For that reason, the occurrence of "four is greater than three" in (e) is extensional and the corresponding occurrence in (a) is not.

I think this proposal is close to the truth, but it is not quite accurate. Consider:

(*) "either snow is white or Mozart wrote music".

(*) doesn’t ascribe truth to either of the propositions associated with the component sentences, but they constitute extensional contexts.

I believe that sentence-connectives are referring terms. Consider "if Mozart wrote music, then somebody wrote music". For reasons given earlier, it seems that the "if…then…" denotes the consequence-relation. Similar remarks apply to "and", "or", "because", "it is not the case", and any other sentence-connective one might think of. Another example might be appropriate. Consider "grass is green, because grass contains chlorophyll." This is true exactly if the propositions *that grass is green* and *that grass contains chlorophyll* are both true and are such that the truth of the second is a consequence of the truth of the first. So "because" can be seen as denoting a function that assigns truth to an ordered pair of propositions exactly if both

are true and the truth of the second is a consequence of the truth of the first. There is no difficulty extending this line of thought to all sentence-connectives.

I would propose that a truth-functional connective is one that assigns a *distribution* of truth-values to a sequence of propositions. If the connective is unary – like "it is false that" or "it is true that" - the sequence consist of one. Consider

(**) "grass is green or snow is white" .

This is true exactly if not both of propositions associated with the component sentences are false. There are four possible truth-value-distributions that a pair of propositions might have: true/true, true/false, false/true, and false/false. (*) is true exactly if one of the first three distributions applies to the propositions in question. So "or" can be seen as denoting the property of having one of those truth-value-distributions. It is clear how this analysis can be extended to any other truth-functional connective. So a truth-functional-connective can be regarded as one that denotes a property had by sequences of propositions: the property of being such that certain truth-value-distributions apply to it.

For reasons outlined earlier, and to be given in full in due course, I think that even *atomic* sentences comprise components that refer to propositions, and that such sentences ascribe truth to propositions. So even where atomic sentences are concerned, truth – and thus a distribution of truth-values – is being ascribed to a proposition.

This line of thought validates our intuitions. We believe, intuitively, that the occurrence of "three is greater than four" in

(a) John believes that three is greater than four"

has the same meaning as its counterparts in:

(e) "it is true that four is greater than three"

and

(g) "four is greater than three".

If the view just outlined is correct, this intuition is to be trusted. What differs from case to case is *not* what "four is greater than three" means, but rather what is being said *about* the thing it means. It means the same thing in each case. But in (a) the property of being believed by John is ascribed to it, whereas in (e) the property of being true is ascribed to it. (Actually, in (e), as we will see, the property of true that it is true is ascribed to it. In (g) the property of being true is ascribed to it. But we can ignore this subtlety for now.) The point is that whether a sentence s occurs truth-functionality has to do *not* with what it means, but with what is being said *about* the thing it means. I think it is absurd to suppose, with Frege, that "four is greater than three" means one thing in (g), a different thing in (e), yet another thing in "it is true that it is true that four is greater than three". The view proposed gives us a way of avoiding that strange view.

§ Let us sum up what we've said in the last two chapters. (LL) cannot be non-question-beggingly used in the context of the Slingshot; and it is almost certainly false. (CR) is almost certainly true. But, ironically, the Slingshot requires that it be true *only* in certain contexts – in so-called extensional contexts. The Slingshot thus requires that there be *non*-extensional contexts: for if (CR) holds without fail, then (SS) becomes false; and if (SS) is false, the Slingshot is immediately voided. The problem is this. Suppose we grant that there are intensional contexts. We cannot define an "intensional" context as one where (CR) doesn't hold; because then (CR) becomes trivial: intersubstitutions of co-referring terms preserves reference – except when it doesn't. And we can't define an "intensional" context as one where a proposition is rigidly denoted: for that would falsely make *all* contexts intensional, given a plausible conception of reference. So if we grant the semantics which the Slingshot needs, then (CR) is stripped of any content, and is therefore useless to the Slingshot. More importantly, any semantics which strips (CR) of content must be rejected. So there is no tenable semantics that will validate the Slingshot.

§ Now it is time to apply the points we've been developing to the Slingshot. We've seen that, if (SS) is false, then the Slingshot immediately fails. (We've also seen some reason to be believe that (SS) *is* false.) In this section, we will suppose, for the sake of argument, that (SS) is true. And we

will see that *even if this is granted*, the Slingshot fails. So we will see that the Slingshot fails on its own terms – it fails even relative to its own semantics.

Chapter 3 The Concept of Logical Equivalence: Predicate-equivalence vs. sentential-equivalence vs. sentence-token equivalence vs. propositional equivalence

So far we've made a *prima facie* case that the Slingshot cannot use (LL). But perhaps there is still hope. Suppose that (LL) could be *independently* motivated – suppose it could be shown that (LL) was true. In that case, perhaps, it *could* be used in the context of a Slingshot argument.

In this chapter, I wish to show that the exact opposite is the case. There are no demonstrative grounds for (LL). And, when we register a couple of key insights of contemporary semantics, there turn out to be no intuitive grounds for it.

So far as I can tell, there are a few different reasons – but none of them is compelling. One is the idea – prevalent in the days of logical positivism, but now generally rejected -- that logically equivalent sentences "say the same thing" and are thus synonymous; and, being synonymous, they must co-refer (in any case, they cannot *differ* in reference).[25]

One problem with this argument is that logically equivalent sentences sometimes *don't* say the same thing – "triangles have three sides", "1+1=2". Another is that, even if it were granted that they always do say the same thing, it isn't clear why they would have to *refer* to anything.

There is, I believe, another reason that LL is sometimes accepted. In his version of the Slingshot, Davidson uses the principle that logically equivalent *singular terms* necessarily co-refer.[26] From this it follows that sentences (if singular terms) must have the same referent if they are logically equivalent. Consider the expressions: "the unique x such that x=Plato and 1+1=2" and "the unique x such that x=Socrates if 1+1=3 and x=Plato if 1+1=2". Being logically equivalent, they must co-refer. So, up to this point, Davidson's principle seems pretty reasonable. When we apply it to logically equivalent sentences, it is natural to conclude that they too must co-refer.

Most versions of the Slingshot do not *overtly* use the assumption that logically equivalent *singular terms* co-refer – they use the assumption that logically equivalent *sentences* co-refer. But I think that part of the motivation for LL is precisely the belief that logically equivalent *singular terms* must co-

refer; LL is thought to be an innocuous extension of that principle.

If this is indeed the reasoning behind LL, then LL is founded on poor reasoning. When you say that "triangles have three sides" and "1+1=2" are "logically equivalent", you are saying that the proposition meant by the one entails, and is entailed by, the proposition meant by the other. But that is *not* what you are saying when we describe *sub*-sentential expressions as logically equivalent. When one says that "the unique x such that x=Socrates if 1+1=3 and x=Plato if 1+1=2" is "logically equivalent" with "the unique x such that x=Plato and 1+1=2", one isn't saying that the *proposition* meant by the one entails the proposition meant by the other. One is making a rather different statement. Fro something to be the referent of "the unique x such that x=Socrates if 1+1=3 and x=Plato if 1+1=2", it must satisfy the open-proposition:

(a) x is a unique x such that x=Socrates if 1+1=3 and x=Plato if 1+1=2" is "logically equivalent.

For something to be referred to by "the unique x such that x=Plato and 1+1=2", that thing must satisfy the function:

(b) x is a unique x such that x=Plato and 1+1=2."

When one says that those two definite descriptions are "logically equivalent", one is saying that, as a matter of logic, if a thing satisfies the condition necessary to be *referred to* by the one description, then it must also satisfy the condition necessary to be *referred to* by the other description. Now when you say that "1+1=2" and "triangles have three sides" are logically equivalent, you are not saying anything about the conditions that things must satisfy to be *referred to* by those expressions; one isn't saying anything about reference. So one is guilty of an equivocation in saying that logically equivalent *sentences* must co-refer on the grounds that "logically equivalent" definite descriptions must co-refer: two different senses of logical equivalence are at work.

Now there is a way to deal with this – and this brings us to what I believe is the third possible motivation for LL. Definite descriptions (plausibly) have both sense and reference. The sense, presumably, is

something like a concept – something like a condition that things may or may not satisfy. The sense of "the king of France" is the concept *unique king of France*, which is not so different from the propositional-function: *x is a unique king of France.* The referent is the thing that satisfies that open-proposition. So the sense is (in effect) a propositional-function (*x is a unique king of* France), and the referent is the thing which satisfies that function. The sense is a concept of the referent.

Given this, one could make the following reasonable claim: the sense of a whole sentence is a proposition; a proposition can be seen as a kind of propositional function – perhaps as a limiting case of such a thing. The referent of a sentence is the thing which *satisfies* that function.

Then one would close the argument by saying: the thing which satisfies such a function is always a truth-value. So propositions are concepts of truth-values. Kaplan[27] once wrote:

[T]he question of the truth value of which a given proposition actually is a concept and the individual of which a given individual concept actually is a concept are…empirical. That is, although a given proposition may actually be a concept of Truth, there are usually other possible states with respect to which it is a concept of Falsehood.

In addition to being concepts of truth-values, propositions are the senses of sentences. So, to conclude this argument, sentences refer to truth-values, just as "the inventor of bifocals" refers to Franklin.

But there is a problem here. If a proposition is a kind of function, then it is a function of zero-argument places. So every true proposition is vacuously satisfied by everything, and every false one by nothing. Thus if we accept that propositions are the senses of sentences, then the *referent* of a proposition is either *everything* or *nothing*, depending on whether it is true or false. But in that case, logically equivalent false sentences don’t co-refer (things that don’t refer at all cannot *co*-refer). Second, the moment we accept the idea that propositions are *concepts*, we are bound to accept that materially equivalent propositions are concepts of *everything* – not just truth-values, but water and peanut-butter and propositions and everything else. So it then becomes a little unclear how one could use the principle that logically equivalent sentences co-refer within the context of an argument purporting to show that sentences refer (exclusively) to truth-values.

There is, perhaps, another way to conceive of propositions; and it might be thought that LL is validated by this other conception. We are generally told that some propositions have different truth-values in different possible worlds. So *Socrates was wise* is true in w1 and false in w2, and so forth. At the same time, we want to say that there is such a thing as *the* proposition *Socrates was wise.* So that proposition is assigned *truth* in w1 and falsity in w2, and so forth. It thus pairs off worlds with truth-values – it is, in effect, a function from worlds to truth-values.

If that is right, then propositions are concepts of *worlds* (or sets of worlds), and not truth-values. Consider the function *x is bald*. We can see this as pairing off individuals with truth-values. Since Socrates is bald, that function pairs off Socrates with the truth-value *true*. Since John Kerry is not bald, it pairs him off with the truth-value *false.* And so on. Now x is bald is a concept *not* of a truth-value, or truth-values; so far as it is a concept of anything, it is a concept of the things *to which* it assigns truth-values. Socrates falls under the concept *x is bald*. So, arguably, *x is bald* could be said to be a concept of Socrates. But *x is bald* is definitely *not* a concept of a truth-value. Concepts are concepts of the things *to which* they assign truth-values, not of truth-values themselves.[28] Now if it is a concept, the proposition *Socrates was wise* is a concept not of truth-values themselves, but of the things *to which* it assigns truth-values, those things being worlds. So *if* propositions are concepts, they are concepts of *worlds* (or perhaps sets of worlds).

Intuitively this makes a lot more sense than saying they are concepts of truth-values. The function *x is bald* is something that is true *of* Socrates and false *of* Kerry. The proposition *Socrates is wise* is true *of* this world and false *of* others. So *Socrates is wise* is something which *describes* this world, and applies to it. Just as *Socrates was bald* describes Socrates, so *Socrates was wise* describes this world, and can thus, without too much artifice, be seen as a concept under which this world falls – as a *concept of this world.* Perhaps one could make some kind of case that it is also a concept of truth-values; but it seems more natural, and consistent with what we already believe about concepts, to see them as concepts of the world (or of sets of worlds).

If I am not mistaken, there is a related problem with the idea that propositions are concepts of truth-values. Consider the concept *unique inventor of bifocals.* This is surely a concept of Franklin. It is a concept of

Franklin because Franklin falls under it – because Franklin has the corresponding property.[29] Roughly, a concept of a thing is a property had by that thing. (It might be more correct to say that a concept is a mode of presentation of a property. But we can attend to that nuance later.)

In light of this, let us consider the relation between propositions and truth-values. Truth and falsity are properties of propositions, not *vice versa*. The proposition *snow is white* falls under the concept *true*; it has that property (or, in any case, the property of which that concept is a mode of presentation). So truth and falsity are concepts of propositions. Propositions are not properties or concepts of truth-values.

Later on I am going to defend an idea found in Lewis (1986): a proposition is (very roughly) a *property* of a world. For a proposition to be true is for the world to be a certain way. *Smith is tall* is true in a world w exactly if w is a certain way – exactly if, to use Kenneth Taylor's phrase, the "quantum" in w is "rippled" the right way. So a proposition can be seen as a *way* that worlds can be; it can thus be seen as a property of worlds. For a proposition to be *true* in a world is for it to be *instantiated* in that world. So truth is instantiatedness. (So I will argue. I should say that, in order to account for the fact that propositions are rich in compositional structure, I am going to have to qualify Lewis' pivotal insight.)

If this is right, then truth is a property of properties: truth is the second-order property of instantiated. (Actually, we will see that truth is a property of *sets* of properties, and is thus a *third*-order property!)

So propositions are properties, and truth is the higher order property of being instantiated. A concept or property of x is of a "higher order" than x. So propositions cannot be concepts, or properties, of the property of being true. So if we plausibly identify the truth-value True with the property of being true, then propositions cannot be concepts of truth-values. So even if we accept Frege's sense-reference semantics, and accept that propositions are the "senses" of sentences, we are bound to conclude that sentences do *not* express concepts of truth-values and *do* express concepts of worlds (or sets of worlds). So relative to Frege's own semantics, if we want to assign denotations to sentences, those denotations must *not* be truth-values, and must be worlds (or sets thereof).

In any case, if the view just outlined is at all close to the mark, then propositions are properties, and therefore concepts, of worlds; and propositions cannot be concepts of truth-values, since truth-values are of the

wrong logical type – they are too high up in the hierarchy.

§ There is another, more interesting problem with (LL). Two pieces of background information must first be given.

The first is the distinction between sentence-types and sentence-tokens. For reasons we considered earlier, it is sentence-tokens that encode propositions. The content of a sentence-type is a function that assigns propositions to its tokens on the basis of context. This is obviously the case with sentences that contain indexicals. The type "I am tired now" doesn't encode a proposition. The content of the type is a function that assigns propositions to its tokens on the basis of facts about the context of utterance – who is doing the speaking, and when the utterance occurs. So some kind of "two-dimensionalism" is needed – the meaning of a sentence-token is one thing, and the meaning of a sentence-type is another.

It might be thought that two-dimensionalism is needed only for a special sub-class of sentences. Sentence-types seem not to have a context-sensitive component. Perhaps where they are concerned, token-meaning and type-meaning coincide.

I think this is a mistake. First of all, every sentence of natural language comprises a tense-marker. Tense-markers seem to be indexicals that refer to times. And even if, technically, they aren't indexicals, their presence always creates a gulf between type- and token-meaning, and thus warrants a two-dimensionalism. So they are *like* indexicals in the respect that is relevant in this context. Consider the sentence-type: "Bob is in the kitchen". A token of that type that occurs at t is true, and a token that occurs at t* is false. That isn't because one and the same proposition is both true and false: we don't need to accept the existence of true-contradictions. What is going on is much more innocent. The one proposition is true iff Bob is in the kitchen at t, and the other is true iff Bob is in the kitchen at t*. Obviously that truth-conditional difference derives form the presence of the case-marker. (The tense-marker is not phonetically realized in all languages. But it is still doing its job.) So tokens of that type encode different propositions. So, in that case, the presence of case-markers warrants a two-dimensionalist approach.

Independently of this, considerations of uniformity warrant our applying two-dimensionalism to *all* sentences, even those that appear resolutely context-insensitive. (Sentences of formal languages, or formal

extensions of natural language, seem context-insensitive. Consider the type "1+1=2".) It is very hard to believe that *in some cases* the property of bearing a proposition is the property of spatio-temporal entities (tokens), while in others it is the property of abstract objects. It seems to me that the phenomenon of indexicality exposes a deep and pervasive fact about linguistic meaning.

In any case, we've seen some reasons to believe that it is sentence-tokens, not sentence-types, that encode propositions. Let us operate on the assumption that this is the case.

The phenomenon of indexicality creates – or, better, *exposes* – a plurality of kinds of "logical equivalence". The propositions like *triangles have three sides* and *1+1=2* are logically equivalent.[30] Sometimes when we say that sentences (or sentence-tokens) are "logically equivalent", we mean that the corresponding propositions are logically equivalent.

But Strawson exposed a different kind of logical equivalence. First of all, Strawson distinguishes tokens and types. He says that a token of "the current U.S. President snores" is true exactly if there is a unique U.S. President x, and x snores. So Strawson would grant that a given *token* of "the current U.S. President snores" is true exactly if a token of "there is a unique U.S. President x, and x snores" is true. But in Strawson's view (as well as mine), a token of the former encodes a proposition of the form *O snores.* If that token occurred today (2005), it would encode the proposition *Bush snores.* If it occurred twenty years ago, it would encode the proposition *Reagan snores.*

Let us focus on the proposition

(*) *Bush snores.*

That proposition doesn't entail, and isn't entailed by, the proposition:

(**) there is a unique U.S. President x, and x snores.

So even though a given token of "the current U.S. President snores" is true exactly if the same is true of a contemporaneous token of "there is a unique U.S. President x, and x snores", the corresponding propositions aren't equivalent at all. There is sentential, but not propositional, equivalence.

Kaplan extended and popularized Strawson's insight. A given token

of "I am here now" must be true, and is thus true exactly if the same is true of a token of "1+1=2". But a token of "I am here now" uttered by Fred (at time t, place p) encodes a contingent proposition (*Fred is in p at t*).

Let us return to Strawson for a moment. How can two sentence-tokens be logically equivalent if they don't encode equivalent propositions. Here is one answer. (Here I am going beyond what Strawson explicitly says.) Intuitively, "the current U.S. President" seems to be an expression that refers to somebody – to Bush, as it happens. But the semantic rule for that expression – the rule that gives it meaning – is obviously *not:*

(i) "the current U.S. President" refers to Bush.

To a first approximation, the right rule is:

If somebody O is a unique president, *then* "the current U.S. President" refers to O.

Actually, a less approximate statement of the right rule is this:

(ii) For any time t, *if* at t somebody O is a unique president, *then* a token at t "the current U.S. President" refers to O. [31]

So given that Bush is President in 2005, a token at that time of "the current U.S. President" refers to Bush.

The semantic rule for:

(iii) "the current U.S. President snores"

is obviously *not*:

"the current U.S. President snores" means: *Bush snores.*

The right semantic rule is (approximately):

(iv) If O is the unique U.S. President at time t, *then* a token at t of "the current U.S. President snores" means: *O snores.*[32]

Given these points, it is immediately obvious how a token of "the current U.S. President snores" could be logically equivalent with a token of "something x is uniquely a U.S. President, and x snores", even though the corresponding propositions are not equivalent.

It is also clear how the position just described elegantly reconciles the plausible view that "the current U.S. president" is a singular term with the fact that the proposition that is *communicated* by a token of "the U.S. president snores" is very different from that communicated by a token of "Bush snores".

Presumably English speakers know the rules which assign meaning to the expressions of their own language. (In any case, they know the basic ones, and at least some of the derivative ones. Otherwise it would be impossible to explain how anyone ever understood what anyone else was saying.) At some level, in some way, English speakers surely know that "snow" means *snow* and "snow is white" means *snow is white.* And surely this point generalizes to other utterances.

Given this, suppose for the sake of argument that (iv) gives the right semantic rule. Any English speaker hearing a token of "the current U.S. president snores" will know that, *if* it is to be true, there must be exactly one U.S. President O, and also that O snores. So an utterance of that sentence will inevitably convey the proposition: *there is exactly one U.S. President O, and O snores.*

There is a difference between hearing the *noise* "the current U.S. president snores", on the one hand, and knowing what it means. A person who doesn't speak English hears the noise but doesn't know the meaning. The gulf between hearing the noise and knowing the meaning is bridged by some kind of background knowledge; this includes (but is not limited to) knowledge of the right semantic rules. Supose that (iv) is the right rule. In that case, one assigns the right meaning to the noise via one's knowledge that *if* there is a unique President O at the time of utterance, then that noise is true exactly if O snores. So to any competent speaker, that utterance will be associated with a certain existence-claim. But that association is *pre*-semantic. The semantics of that noise is the proposition *Bush snores.* But to assign the right proposition to that noise, you have to do some work: you compute knowledge of semantics on the basis of semantic rules (and on the basis of other kinds of knowledge as well: a topic we will bracket for now).

We might describe this work as pre-semantic – it is the work that is pre-requisite to a knowledge of semantics. Supposing that (iv) is the right rule, it actually *follows* that "the current U.S. president snores" will communicate an existence-claim – the very claim that Russell saw as its literal meaning. But (iv) doesn't have the counter-intuitive (and, I will argue, demonstrably false) consequence that "the current U.S. President" is anything other than what it seems to be.

Notice that (iv) gives a general rule. It doesn't talk about this or that specific sentence-token. It talks about the corresponding *type.* It says that given *any* token of that *type*, the meaning of that token is such and such. So (iv) enables one to commute token-meaning by giving type-meaning. We thus see how, to an extent, pre-semantically imparted information coincides with *type*-meaning. (But , as we'll see, much pre-semantic information is of a different, not specifically linguistic kind.)

The semantic rule for "George Bush snores" is obviously *not* (iv): it is (leaving out irrelevant facts about the derivation –tree):

(v) A token of "Bush Snores" means *Bush snores.*

When a person hears a token of "Bush snores", the pre-semantic work he does to assign it the right meaning don't say anything about the number or existence of U.S. Presidents. So what is pre-semantically implicated by a token of "Bush snores" will not remotely like what is thus communicated by a token of "the U.S. President snores".

An exactly similar story explains why tokens of "that man over there is tired" and "Bob is tired" and "you are tired" will tend to communicate very different propositions, even if one and the same person is referred to in all three cases.

To sum up, there is nothing strange about Strawson's contention that sentence-logic may diverge from proposition-logic. On the contrary, granted the assumption that speakers of a language have some minimal awareness of the semantic rules characteristic of that language, Strawson's contention reconciles our intuition that definite descriptions are singular terms with the facts about the cognitive significance of expressions containing definite descriptions.

But right now, we are not worried about the correct semantics of definite descriptions, or about the nature of reference. We are going to focus

on the facts about logical equivalence brought to light by Strawson's groundbreaking work – his distinction between type- and token-semantics (he uses the term "utterances", instead of "tokens"), and his distinction between expression-logic and proposition-logic.

Let T and T* be two logically equivalent sentences-tokens, encoding propositions P and P*. Suppose that P and P* are *not* logically equivalent. (So T and T* might be contemporaneous tokens of "I am here now", uttered by Fred in Paris in 2005, and "Fred is in Paris in 2005" uttered by someone else.) Advocates of (LL) presumably don't have this sort of situation in mind. Even if -- along with Stalnaker, Lewis, and the logical positivists – we *identify* logically equivalent propositions, we cannot do that with the propositions involved here. Also, there are no intuitive or demonstrative grounds for saying that P and P* are concepts of the same thing. A proposition is a description, and therefore a way of presenting, the world. So, as we discussed earlier, it certainly seems that a proposition is a concept of a world. P is true in some worlds where P* is not true, and vice versa. So P and P* are concepts of different worlds and thus are not concepts of the same thing.

The advocate of the Slingshot could circumvent this problem. In some cases, two sentence-tokens are logically equivalent in the sense that they encode logically equivalent propositions. (So a token of "triangles have three sides" is P-logically equivalent with a token of "squares have four sides".) Let us describe such pairs as P-logically equivalent. instead of using (LL), he could use the principle that:

(PLL) P-logically equivalent sentences (or sentence-tokens) co-refer.

The problem is that the very things that (PLL) presupposes make it hard to believe that sentence-tokens refer to truth-values.

(PPL) is motivated by, and therefore presupposes, the existence of a distinction between expression-logic and proposition-logic. What I am calling sentence-logic corresponds very closely to Fregean sense. As a result, once we grant the mere existence of a distinction expression-logic and proposition-logic, it becomes impossible not to grant that sentence-tokens refer to the propositions they encode.

Let T be a token of "I am here now", uttered by Fred at time t in place p. Let T* be a token of "Fred is in place p and time t." Given what

Kaplan (1989) says, we have good reason to believe that T and T* both have the same proposition for their literal meaning (*Fred is in place p and time t*). But there is no doubt that T and T* *communicate* very different propositions. So – whatever their literal meanings might be – there is no doubt that they have dramatically different *cognitive* values.

What would a Fregean say here? He would say T and T* have different *senses.* In any case, *if* Frege's sense-reference semantics applies anywhere, it applies here. What is communicated by T is trivial; what is communicated by T* is non-trivial. The sense of the one is trivial; the sense of the other is not.

Let me be as clear as possible on what I am saying. Kaplan argues that T and T* encode the very same proposition (*Fred is in p at t*). That is well and good (I happen to think it is correct). But be that as it may, T and T* certainly have very different *cognitive* values. So they have very different *senses* (sense just *is* cognitive value). So here we have a text-book example of a case where Frege's notion of sense applies. If we wish to remain loyal to a position that is even remotely Fregean, we must say that T and T* differ in sense.

What I am now going to argue is that the two senses in question pick out *propositions* (the same in each case). And I will also argue that the difference in sense derives, not from semantics, but from *pre*-semantics – from the work one must do to compute literal meaning. The position we end up with is this: suppose we wish to say that T and T* are referring terms. In that case, when we apply Frege's sense-reference distinction to the case at hand, we must conclude that T and T* denote the propositions they encode, and do *not* denote truth-values.

How does an English-speaker hearing T know what it means? He does so on the basis of his knowledge of the relevant semantic rules. What are those rules?

If O tokens a sentence of the form *…I…*, then that token means (or is true exactly if) O has…x… If Smith tokens a sentence of the form "…I…", then that token is true exactly if Smith has…x…If Smith says "I am thirsty", that token is correct exactly if Smith is thirsty.

If, at place P, a sentence of the form *…here…* is tokened, that token is true exactly if P has…x…So if the sentence "Smith is here" is tokened in Paris, then that token is true exactly if Smith is in Paris.[33] A similar rule applies to sentences of the form "…now…"

The rule that assigns meaning to T is a derived one, comprising the rules just mentioned (and many others). That rule is (very approximately):

(^) If "I am here now" is tokened by O in place p at time t, then that token is true exactly if O is in place p at time t.

(^) entails that, for a token of "I am here now" to be true, the speaker must be where he is, when he is there. Such a token is true as long as the speaker doesn't manage to simultaneously be, and not be, in a single place. That is why that *token* is "logically true", even though the corresponding proposition is not.

A Fregean would say that the "sense" of T is trivial. We've seen that what is trivial is the information *through which* one grasps the proposition literally meant by T. So the sense just mentioned picks out a *proposition*. The thing literally meant by T is some proposition. The "sense" of T is the information which enables one to single out that proposition. So the sense of T is something that singles out a proposition. The Fregean view is that the referent of an expression is the thing which the sense of that expression singles out. So if we insist on saying that T has a denotation, then application of the sense-reference distinction forces us to say that the denotation is the proposition meant by T. The sense is pre-semantics; the referent is semantics (literal meaning, the proposition literally meant).

(PLL) embodies a recognition of the fact that expression-logic may diverge from proposition-logic. But that very fact exposes Fregean "sense" for what it is: *pre*-semantics, as opposed to semantics proper. The "sense" of T – the thing that makes the *expression* trivial – is the information by means of which one identifies the proposition meant by T. So if we insist on saying that it has a referent, that referent coincides with the proposition that it has for its literal meaning.

§ Let us now consider Church's (1943[34]) argument for the thesis that sentences refer to truth-values.

Let us assume that:

(LL) logically equivalent sentences co-refer.
(CR) Intersubstituting co-referring terms preserves reference.

Let S and S* be any two non-analytic sentence-tokens that have the same truth-value.

(1) S
(2) The class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x).
(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).
(4) S*.

(1) and (2) are logically equivalent. So (1) and (2) co-refer. (3) is what results when a referring term in (2) is replaced with a co-referring term. So (2) and (3) co-refer. (3) and (4) are logically equivalent. So, by (LL), (3) and (4) co-refer. Thus, (1) and (4) co-refer. What we've just said about (1) and (4) can be said (mutatis mutandis) about any two sentences that are alike in truth-value. So any two sentence-tokens, alike in truth-value, co-refer. If this is right, then it is immediately obvious that sentences do *not* refer to the propositions they have for their meanings. And it is also strongly suggested that sentences *do* refer to their truth-values.

Let us now evaluate this argument. One problem lies with (LL). As we've seen, use of (LL) in this context is question-begging. (LL) nearly enough embodies the assumption that sentences *don't* refer to propositions, and is thus question-begging. It seems pretty clear that necessarily equivalent sentences can express very different propositions: "triangles have three sides", "1+1=2", "there are infinitely many primes". So *if* sentences do refer to propositions, then logically equivalent sentences *don't* co-refer. Thus *if* (LL) is to be of any use to somebody trying to argue that sentences refer to truth-values, that person must *assume* at the outset that sentences don't refer to propositions. But this is not something that can be assumed in such a context. So (LL) is not a permissible assumption in this context.

Also, a case can be made that the use of (CR) in the context of Church's argument is question-begging. As we just saw, someone who is arguing that sentences refer to truth-values cannot presuppose that they *don't* refer to propositions. And as we saw earlier, the Slingshot presupposes that the meaning of a sentence, i.e. the proposition it expresses, *can* be changed by intersubstituting co-referring terms. But if sentences refer to propositions,

then such substitutions *do* change the referent of a sentence. So if (CR) is to be used in this context, it must be assumed that sentences do not refer to propositions. But this assumption is question-begging.

The counter-response would be this: "It is clear *on independent grounds* that (CR) holds in extensional contexts; and the intersubstitutions involved in the Slingshot occur in extensional contexts". But we've seen that there is no definition of "extensional" relative to which this reply works. If "extensional" is defined one way, then (CR) is stripped of any content, and all that is left is the innocuous triviality: "intersubstituting co-referring terms preserves reference, except when it doesn't". If "extensional" is defined another way, it is left open whether intersubstituting co-referring terms change sentence-reference.

To sum up, neither of the assumptions of Church (1943) can be legitimately used in that context.

Let us now consider Church's 1956 argument:

(i) Sir Walter Scott is Sir Walter Scott.

(ii) Sir Walter Scott is the author of Waverly.

(iii) Sir Walter Scott is the man who wrote twenty-nine Waverly novels altogether.

(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.

(v) The number of counties in Utah is twenty-nine.

This argument uses two principles. One is (CR). The other is:

(NS) Expressions that are synonymous, or *almost* synonymous, co-refer.

Here is the argument proper. (ii) is what results when a referring term in (i) is replaced with a co-referring term. (iii) is what results when a referring term in (ii) is replaced with a co-referring term. So all of (i)-(iii) co-refer. (iii) and (iv) are "nearly synonymous" and therefore co-refer. (v) is what results when a referring term in (iv) is replaced with a co-referring term, so (v) and (iv) co-refer. Thus (v) co-refers with (i). Of course, (v) and (i) have very different *propositions* for their meanings. So we must not say that sentences

refer to propositions. So what do they refer to? Intersubstitutions of co-referring terms cannot change reference. So reference has been preserved during the journey from (i) to (v), and so has truth-value. Thus it is natural to identify sentence-truth-value with sentence-reference. (Of course, there are other semantically significant properties that both (i) and (v) have in common; for example, they are both in English. That said, *if* we accept Church's assumptions, then the argument is certainly suggestive: it suggests that truth-value – and not, say, the property of being in English – is what (v) and (i) both refer to.)

We've already seen why (CR) cannot be used in the context of an argument purporting to show that sentences refer to truth-values. So that right there vitiates Church's argument.

Let us focus on (NS). There are many problems with this assumption.[35] Why should *nearly* synonymous expressions co-refer? Compare "the most famous hair-stylist", which refers to (let us say) Vidal Sassoon, and "the most famous barber", which refers to someone else.

Of course (iii) and (iv) are (almost: see below) logically equivalent. So if we accept (LL), then they co-refer. But we've already seen the problems with (LL): it cannot be used in this context. Also, (iii) and (iv) are, at least arguably, different propositions. So *if* sentences refer to propositions, then (iii) and (iv) don't co-refer. So to ensure that (iii) and (iv) co-refer, we'd have to assume that sentences *don't* refer to propositions; and, as noted earlier, this assumption cannot be made in this context.

There is a minor point. (iii) and (iv) are not *quite* logically equivalent. (iii) implies that the Waverly novels were written by a man, and also by exactly one thing: (iii) talks about *the* man who wrote the Waverly novels, and also the *man* who wrote those novels. (iv) does not say or even imply that they were written by a man. And it is unclear whether it implies that Scott *uniquely* wrote them. (It seems not to imply this, at least not strictly.)[36]

It would be possible to deal with this. We change

(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.

into

(iv*) The number n, such that some man identical with Sir Walter Scott uniquely wrote N Waverly novels, is twenty-nine.

But, even with this change, it seems to me that (iv*) and (iii) are not *quite* synonymous. It would be an exercise in logic, albeit an easy one, to verify their equivalence. And where one has to *prove* equivalence, it is doubtful whether there can be synonymy.[37] So (iv*) and (iii) are at best *nearly* synonymous. But we've already seen that near synonymy doesn't guarantee co-reference.

To sum up, neither (CR) not (NS) is an assumption that can be made in this context. (NS) is dubious; and (CR) is false unless duly restricted, but thus restricted it cannot be non-question-beggingly used here.

Gödel's Slingshot

Gödel gave an argument purporting to show that, given certain assumptions, it can be proven that all true sentences have the same "signification". It isn't clear whether, by "signification", Gödel meant "denotation" or "sense".[38] But either way the argument fails. Here we needn't go through the whole argument. We need only state the assumptions it uses.

(Terminological point: *s(a)*: the signification of *a*. The asterices are quasi-quotes.)

G1: if *a=b* is true, then *s(a)*=*s(b)*.
G2: There is a function f such that for every a, b such that *a=b* is well-formed, *s(a=b)*=f(s(a), s(b)).
G3. s (*…a…*)=s(* the unique x such that x=a and…a…)*).

G4. *Fa* is logically equivalent with *a=the unique thing x such that x=a and Fx.*

Gödel's argument also assumes that definite descriptions are singular terms.

Here we must remember one of the points we made when we were

discussing this argument earlier. Consider G4. If we *reject* (SS), then (G4) is simply false.

Consider the sentence: "the unique x such that x=Socrates and x is bald is identical with Socrates". If (SS) is false, then that sentence encodes the proposition *Socrates is identical with Socrates.* In that case, it is *not* logically equivalent with "Socrates is bald". So Gödel's Slingshot presuppose the truth of SS.

So, for the sake of argument, let us suppose that (SS) is right. In that case, the other premises are inconsistent with another. First of all, the term "signification" is ambiguous between "meaning" and "denotation". If by "signification" is meant "meaning", then G1 is straightforwardly false, if one accepts (SS). "The first post-master general=the inventor of bifocals" is true. But, given (SS), the *meaning* of "the first post-master general" is not identical with the *meaning* of "the inventor of bifocals". So, since we are supposing (SS) to be true, we must take "signification" to mean "denotation". If we do that, then G1 is indeed true.

But then G2 becomes either false or question-begging. Consider the sentence "the first post-master general=the inventor of bifocals". It is by no means clear whether it *has* a denotation. If that is the case, then there is no function that *assigns* it a denotation (at least no function that assigns it the denotation that it *actually* has). So if G2 is to hold, it must be assumed that sentences *do* denote.

But even this assumption is not enough to make G2 be true. G2 is to the effect that the denotation assigned to "the first post-master general=the inventor of bifocals" is a function only of the *denotations* of the components of that sentence, and *not* of their meanings. But what if sentences denote propositions? In that case, "the first post-master general=the inventor of bifocals" will *not* refer to the same thing as "the first post-master general= the first post-master general"; and the referent of a sentence will be a function of the *meanings* of its components, and not (just) of their referents. So if G2 is to hold, it must be assumed at the outset that the denotation of a sentence is *not* a proposition. But, as we've seen time and again, this is not something that can be assumed in this context. So G2 embodies just the kind of begging of the question that we saw embodied by (LL) in Church (1943).

Let us now turn to G3. This says that "Socrates is bald" and "the unique x such that x=Socrates is bald" co-refer. But why should they co-refer? If (SS) is correct – and it must be for Gödel's argument to go through -

- those sentences encode different propositions. The concept of identity (and uniqueness) is a constituent of the one proposition, but not (at least not obviously) of the other. So even though those sentences are obviously equivalent, it is still an exercise in logic to prove that equivalence; and, as we said, where equivalence has to be *proven*, we are probably not dealing with sentences meaning *precisely* the same proposition. So those two sentences probably have different propositions for their meanings. In that case, *if* sentences refer to propositions, then "Socrates is bald" refers to a *different* proposition from "the unique x such that x=Socrates is bald". So if G3 is to be correct, it must be assumed that sentences *do not* refer to propositions. But, as we've noted, this is something that cannot be assumed in this context.

To sum up: if G1 is to hold, then we must take "signification" to mean "denotation". But taken that way, G2 and G3 become either question-begging or false (or both, if sentences *do* denote propositions) .

Another failed argument concerning sentence-reference

Frege has another argument for the view that sentences refer to truth-values. Reference is invariant with respect to intersubstitutions of co-referring terms. "The father of Cicero" has the same referent as "the father of Tully". (In this context, to help motivate Frege's point, let us assume that definite descriptions are singular terms.) In some contexts – so-called truth-functional ones - truth-value is invariant with respect to intersubstitutions of materially equivalent sentence-tokens: "coal is black and grass is green" has the same truth-value as "snow is white and grass is green". So the facts about reference-preservation seem to parallel those about truth-value-preservation, suggesting that sentence-token-reference is truth-value.

There are two problems with this argument. As Frege well knew, truth-value is *not* invariant with respect to intersubstitutions of materially equivalent sentence-tokens in *non*-truth-functional contexts. "Fred believes that snow is white" is true, but "Fred believes that there are more reals than rationals" is false. It is only in *truth-functional* contexts that such intersubstitutions work. But a "truth-functional" context is *by definition* one where such intersubstitutions work. An operator is truth-functional exactly if, for any sentence s falling in its scope, replacing s with a materially equivalent sentence doesn't affect the truth-value of the host-sentence. So Frege is indeed right that truth-value is preserved by intersubstitutions of material

equivalents in truth-functional contexts. But that point amounts to: *truth-value is preserved by intersubstitutions of material equivalents in contexts where it preserves truth-value* – or, more simply: *such intersubstitutions preserve truth-value – except when they don't.* So facts about truth-value preservation parallel those about reference-preservation – except when they don't.

The truth is that such intersubstitutions tend *not* to preserve truth-value. The only connectives that are truth-functional are "not", "and", "or", "it is true that" (and those that can be manufactured on their basis – e.g. "it is true that it is true…"). Causal, temporal, moral, modal, epistemic, and explanatory connectives are *not* truth-functional. Replacing "Kennedy died" in "Kennedy died because/after Kennedy was shot" with a materially equivalent sentence does *not* preserve truth-value: "Lincoln died because/after Kennedy was shot" is false. So it is more the exception than the rule for intersubstitutions of material equivalents to preserve truth-value; and the fact that truth-value is thus preserved in "truth-functional" contexts is an innocuous tautology.

But even if we leave aside this last point, Frege's argument is still less than probative. For the sake of argument, suppose that *all* contexts were truth-functional and that, consequently, truth-value was *always* preserved under intersubstitutions of material equivalents. Even so, an argument exactly parallel to Frege's suggests that sentence-tokens refer to *propositions*. Let S be any compound sentence-token, whose constituent sentence-tokens are s_1… s_n. For any i, replacing s_i with a sentence expressing the same proposition will result in a compound sentence having the same proposition as S for its meaning. "Snow is white and Frank is a foe" encodes the same proposition as "snow is white and Frank is an enemy." So the facts about *proposition*-preservation parallel those about reference-preservation no less than do the facts about truth-value preservation -- more so when we take into account non-truth-functional contexts. Incidentally, this supports the view that sentences refer to propositions. For it holds *without exception* that if s is a sentence that is a proper part of another sentence S, then replacing s with a *synonymous* sentence s* will result in a host sentence S* that means the same thing as S. The corresponding point about truth-values holds only in the (comparatively rare) cases where s occurs truth-functionally.

Chapter 4 Do Quantifiers refer to functions?

Our consideration of the Slingshot is not yet complete; for there is a relevant subtlety that we have yet to consider. Some background will help to identify that subtlety.

Russell held that definite descriptions are quantifiers. And, in Russell's view, this meant that they don't *refer* to anything. Of course, if Russell is right, and definite descriptions don't refer to anything, then the Slingshot fails. The Slingshot assumes that definite descriptions can *co*-refer. If definite descriptions don't refer to anything, then no two of them refer to the same thing, and no definite description refers to the same thing as a proper name. So no two definite descriptions *co*-refer, and no definite description *co*-refers with a proper name. And that would immediately void every form of the Slingshot.[39]

Our primary argument against the Slingshot did *not* presuppose the truth of Russell's theory of descriptions. That argument had this form. Either that theory is correct or it isn't. If it is correct, then the Slingshot fails (as we just saw). If it is false, then definite descriptions are singular terms, and the Slingshot fails (for entirely different reasons).

But Richard Montague powerfully argued for a thesis that, if correct, apparently shows that argument to embody a fallacy. In this section, I will identify that thesis and show why, even if it is correct, in no way threatens our argument.

Montague's point is this. Quantifiers *can* be regarded as referring terms. This must be understood aright. Russell and Frege showed conclusively that "a man" does not refer to some mysterious "ambiguous" man; and they showed the same thing (mutatis mutandis) of "no man", "every man", and so on. "No man" doesn't refer to some kind of blank man. It doesn't refer to *any* man. Montague agrees with this unreservedly.

Here is where Montague and Russell disagree. Russell took these facts about "a man", and other quantifiers, to show that they don't refer *at all.* Montague argued that quantifiers *do* refer. They don't refer to individuals, but they *do* refer to functions.

Consider the concept *x is bald.* This is true for certain objects (e.g.

Socrates) and false for other others (e.g. Lincoln). So this concept can be seen as a function that pairs off individuals with truth-values. It pairs off an individual x with the truth-value *true* if x is bald and pairs off x with the truth-value *false* if x is not bald. It is thus natural to regard the predicate "bald" as denoting such a function from individuals to truth-values. This is Frege's analysis.

It is also natural to see "bald" as denoting the *property* of baldness. But this view is compatible with the view just described, as we will see in a moment.

Now consider the quantifier "for some x". For some concepts C, *for some x, Cx* is true. For example, "for some x, x is a bald man" is true. For other concepts C, *for some x, Cx* is false. For example, "for some x, x is a square circle" is false. Thus "for some x" can be seen as denoting a function from *concepts* to truth-values. That function assigns truth to a concept C if that concept is instantiated and falsity if C is not instantiated.

So "for some x" can be seen as denoting a second-order function: a function from concepts (functions from objects to truth-values) to truth-values.

Every other quantifier is amenable to an exactly similar treatment. Consider the quantifier "for every x". For any concept C, *for every x, Cx* is true exactly if C is instantiated by everything.[40] Thus "for every x" can be seen as denoting a function that assigns truth to a concept exactly if that concept is universally instantiated. By similar reasoning, "for no x" can be seen as denoting a function that assigns truth to a class C exactly if nothing instantiates C. There is no difficulty extending this analysis to complex quantifiers like "for some bald man x", and "for most birds x".

Let us now return to definite descriptions. Suppose that Russell's theory is correct. In that case "the king of France is bald" is true exactly if there is exactly one king of France x, and x is bald. In general, for any property phi, *the phi has psi * is true exactly if there is exactly one phi x, and x has psi. So, by reasoning analogous to that given a moment ago, we can see *the phi* as denoting a function that assigns truth to a class C exactly if two conditions are met: first, there is exactly one object having phi and, second, any subject object falls under C. For any concept C, *the king of France falls under C* is true exactly if, first, there is exactly one king of France x and, second, x falls under C. So "the king of France" denotes a function that assigns truth to C exactly if both those conditions are met.

To sum up, if Montague is right, then Russell's theory *is* consistent with the idea that definite descriptions are referring terms. So far, we have assumed otherwise. So if our arguments are to work, they must accommodate the possibility that Montague's position is correct. For the record, I think that Montague *is* right.

Fortunately, none of this to any degree threatens anything we've said. For the sake of argument, suppose that Russell is right about definite descriptions; and suppose that Montague is right about quantifiers. In that case, "the inventor of bifocals" denotes a function F that assigns truth to a class C exactly if two conditions are satisfied: first, there is exactly one inventor of bifocals; second, any inventor of bifocals falls into C. And "the first post-master general" denotes a function F* that assigns truth to a class C exactly if two conditions are satisfied: first, there is exactly one post-master general; second, any post-master general falls into C.

F and F* are different functions. There are possible worlds where F assigns truth to Franklin and F* assigns falsity to Franklin. So if definite descriptions denote functions, then "the inventor of bifocals" and "the first post-master general" denote *different* functions.

By exactly similar reasoning, if definite descriptions denote functions, then for any two sentences S and S*, "the class of all x such that (x=x)" denotes a *different* function from each of "the class of all x such that (x=x and S)" and "the class of all x such that (x=x and S*)".

Given this, consider the sequence of sentences involved in Church 1943 and Davidson 1967:

(1) S

(2) The class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x).

(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).

(4) S*.

(3) is supposedly what results when a referring term in (2) is replaced with a co-referring term. The terms in question are "the class of all things x such that (x=x and S)" and "the class of all things x such that (x=x

and S*)." So (2) and (3) are supposed to co-refer by (CR). But supposing that definite descriptions denote functions, those two definite descriptions denote *different* functions and therefore do not co-refer.

So *if* (2) and (3) are to co-refer, definite descriptions must *not* be quantifiers: it doesn't matter whether (like Montague) we think of quantifiers as denoting functions or (like Russell) as denoting nothing. Definite descriptions must be singular terms if (2) and (3) are to co-refer. But we've already seen it won't help the Slingshot if they are singular terms.

Let us now consider the sequence of sentences involved in Church (1956):

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is the author of Waverly.
(iii) Sir Walter Scott is the man who wrote twenty-nine Waverly novels altogether.
(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) The number of counties in Utah is twenty-nine.

(ii) is supposed to be what results when a referring term in (i), namely "Sir Walter Scott", is replaced with a co-referring term, namely "the author of Waverly". So (i) and (ii) are supposed to co-refer by (CR). (iii) and (ii) are supposed to co-refer for the same reason.

Supposing that definite descriptions denote functions, then "Scott" obviously doesn't co-refer with "the author of Waverly". "Scott" refers to a person, not a function. And "the author of Waverly" would not co-refer with "the man who wrote twenty-nine Waverly novels altogether." Suppose that definite descriptions denote functions. Let F be the function denoted by "the author of Waverly". Let F* be the function denoted by "the man who wrote twenty-nine Waverly novels altogether." F assigns truth to a class C exactly if two conditions are met: first, there is exactly one author of Waverly; second any author of Waverly falls into C. F* assigns truth to a class C exactly if two conditions are met: first, there is exactly one author of twenty-nine Waverly novels; second, any such author falls into C. There are worlds where F assigns truth to a given class and F* assigns falsity to that same class. So F and F* are different functions. Imagine a world w where the following holds. In w, Scott is bald. And in w, Scott wrote Waverly, but nobody wrote twenty-

nine Waverly novels. (Scott decided to stop writing Waverly novels after writing the first two.) In w, F assigns truth to the concept *x is bald* (and also to various other concepts: *x is identical with Scott, x is not a round square*). But, in w, F* assigns falsity to *every* class. After all, in w, *nothing* wrote twenty-nine Waverly novels altogether; so in w there is no class C satisfying the two conditions mentioned a moment ago. So F and F* are different functions; therefore, if quantifiers denote functions, then the two definite descriptions being discussed do not co-refer.

So *if* (i)- (iii) are to co-refer, definite descriptions must *not* be quantifiers: it doesn't matter whether (like Montague) we think of quantifiers as denoting functions or (like Russell) as denoting nothing. Definite descriptions must be singular terms if (i)- (iii) are to co-refer. But we've already seen it won't help the Slingshot if they are singular terms.

§ There is a technicality we must consider. According to one point of view, a function is just a set of ordered pairs. So the function *F(x)=x+1* is the set of ordered pairs <0,1>, <1,2>, and so on. From this viewpoint, the function denoted by "F(x)=x+1" would be *identical* with the function denoted by "F(x)=x+(4-(0+1))". Both expressions are associated with the same set of ordered pairs. It doesn't matter that they express that set in different ways. Sometimes this is referred to as the "extensionalist" view of functions.

Given this, there might be a way of reviving the position argued against in the previous section. For the sake of argument, suppose that quantifiers denote functions and that definite descriptions are quantifiers. It might be said that "the inventor of bifocals" *does* denote the same function as "the first post master general". After all, those two expressions are associated with the same set of ordered pairs. Each assigns truth to the concept *identical with Benjamin Franklin*, falsity to *identical with Jefferson*, truth to *witty polymath*, falsity to *brutish moron*, and so on. So, relative to the conception of functions described a moment ago, those two definite descriptions are associated with *the same* function. So, contrary to what we argued, those expressions *do* refer to the same function: they do co-refer.

This reasoning involves a major non-sequitur. But it does force us to make one aspect of our analysis more explicit.

Let us start with the quantifier "for some x". We will then map what we say about this quantifier onto definite descriptions.

In this world,

(*) "for some x, x is a man over 9ft tall"

is false. So in this world, "for some x" assigns falsity to the class of men over 9ft tall. In some other world, (*) is true. So in that world, "for some x" assigns truth to the class of men over 9ft tall.

Given this, it is tempting to say: "for some x" is associated with one set of ordered pairs in w_1, a different set of ordered pairs in w_2, and yet it has the same semantics in both worlds. In w_1, it is associated with the pair <men over 9ft tall, falsity>; in w_2 it is associated with the pair <men over 9ft tall, falsity>; yet in both worlds, it denotes the same function.

But we must not say that. A given function cannot be associated with *different* sets of ordered pairs. A function determines a *single* set of ordered pairs: different sets of ordered pairs, different functions. If we say that "for some x" has the same semantics in different worlds, then we are saying that it denotes the *same* function, and is thus associated with the same set of ordered pairs, in different worlds. So it is a contradiction in terms to say that "for some x" has the same semantics in worlds where it is associated with different ordered pairs.

We need to reconcile the fact that "for some x" can have the same semantics, and thus be associated with the same ordered pairs, in both w_1 and w_2, and yet assign falsity to the class of men over 9ft tall in w_1 and truth to the class of men over 9ft tall in w_2. [41]

This problem is not hard to deal with. Let C_1 be the class of men over 9ft tall in w_1, and let C_2 be the class of men over 9ft tall in w_2. "For some x" does not have *the* class of men over 9ft tall for an argument. There is no such class. There is the class of men over 9ft tall in w_1, the class of men over 9ft tall in w_2, and so on. In other words, there is C_1, C_2, and so on. "For some x" assigns falsity to C_1, truth to C_2, and so on.

So "for some x" denotes a function that comprises the ordered pair< C_1, falsity> , <C_2, truth>, and so on. In every world, "for some x" denotes a function comprising those very ordered pairs. Thus in every world, "for some x" is associated with the very same ordered pairs, and thus doesn’t vary in meaning from world to world.

If definite descriptions denote functions, these points apply squarely to them. "The inventor of bifocals" does not denote a function that assigns

truth to *the* class of snorers: there is no such class. It assigns truth to the class of snorers in w_1, falsity to the class of snorers in w_2, and so on. Let S_1 be the class of snorers in w_1, let S_2 be the class of snorers in w_2, and so on. So the set of ordered pairs associated with "the inventor of bifocals" doesn't vary from world to world: in *each* world that set is <truth, S_1>, <falsity, S_2>, and so on.

For exactly similar reasons, "the first post-master general" isn't associated with *one* set of ordered pairs in w_1 and a different pair in w_2.

For obvious reasons, the set of ordered pairs associated with the one expression will be different from that associated with the others. There are possible worlds where a snorer uniquely invents bifocals, and where a non-snorer is a unique first post-master general. Suppose that w_{76} is that world, and let S_{76} be the class of snorers in that world. In that case, the set of ordered pairs associated with "the first post-master general" includes <falsity, S_{76}>, whereas the set associated with "the inventor of bifocals" does *not* include that pair (but rather includes the pair <falsity, S_{76}>. So *even* if we identify functions with sets of ordered pairs – even if one has a purely "extensionalist" view of what functions are -- the set of ordered pairs associated with the one definite description will *not* be identical with the set of ordered pairs associated with the other. Thus the view proposed at the beginning of this section does not succeed.

Chapter 5 Why sentences refer to propositions: a preliminary argument

We've considered some arguments that are meant to show that sentences refer to truth-values. We've seen that those arguments fail. But that doesn't mean that sentences *don't* refer to truth-values. It means only that the arguments just mentioned fail to establish that they do.

The first question we must ask is this: What is reference? What is it for an expression E to refer to an object x?

Philosophers have said a number of things about sentence-reference: sentences refer to truth-values, to propositions, to facts, to sets of possible facts (or situations), or to nothing at all. But it has never been said what

exactly it would be for a sentence to "refer" to a truth-value or a fact. In fact, what Frege said about the concept of number, and its relation to mathematics, holds equally of the concept of reference and its relation to semantics. In Frege's time, no one had bothered to ask the question "what is a number?" In our time, no one has ever said: "this is what reference is; this is what it is for 'Socrates' to refer to Socrates; and the relation that holds between those two things is identical with the relation that holds between sentences and truth-values (or facts or situations…)"

In this chapter, I want to propose an answer to the question: "what is reference?" I will start by *stating* the proposal I wish to defend.

(*) An expression E refers to an object O exactly if, in virtue of having the form "…O…", a sentence-token encodes a proposition that has O as a constituent.

I do not expect (*) to be at all convincing as it stands. First, we must say why this view prevails against the opposing Fregean view. The Fregean view is that, when "Socrates" occurs in a sentence, the corresponding proposition has for a constituent, not Socrates himself, but rather a *sense* of Socrates (a concept that singles out Socrates). Second, we must say why our view prevails against the popular (though vigorously criticized) "causal theory of reference": E refers to O exactly if tokens of E stand in a certain causal relation to O. Finally, we must make it clear what it means to say that an object is a *constituent* of a proposition. This last problem is far and away the most difficult one.

In this chapter, I would like to give the motivation for (*), and will also say why, in my view, it prevails against the Fregean view. I will discuss the other problems in later chapters.

§ Let us start with some platitudes. We know that "Socrates" refers to Socrates. (Actually, it is *tokens* of "Socrates" that refer to Socrates. But to expedite discussion, let us idealize away from this nicety.) What does this involve? If you say "Socrates was bald", you are attributing baldness to Socrates. If you say "Plato punched Socrates", you are saying of Socrates that Plato punched him. Suppose that "Socrates was bald" did *not* attribute baldness to Socrates. Suppose that, in general, *Socrates has phi* did *not* attribute phi to Socrates, but that otherwise English semantic rules were

unchanged. In that case, "Socrates" wouldn't refer to Socrates at all.

Here is what we have so far. It is a datum that "Socrates" refers to Socrates. It is also a datum that, for any phi, if a sentence has the form *Socrates has phi*, then that sentence attributes phi to Socrates. Surely these facts are related.

Let us move on. Suppose that "Socrates was bald" meant *Charlemagne was bald*, but that otherwise the rules of English semantics were unchanged. In that case, "Socrates" would presumably refer to Charlemagne, at least in that particular context. Suppose that, for any phi, tokens of *Socrates had phi* attributed phi to Charlemagne. In that case, presumably, "Socrates" would refer to Charlemagne, not to Socrates.

Here is what we have so far. If *Socrates has phi* did *not* attribute phi to Socrates, then "Socrates" would not refer to Socrates. So if "Socrates" is to refer to Socrates, it is *necessary* that *Socrates has phi* attribute phi to Socrates.

This condition is also *sufficient.* If sentences of the form *Socrates has phi* meant that *Charlemagne* has phi, then (holding constant all the remaining semantic rules of English) "Socrates" would refer to Charlemagne. In general, for any object O, if *Socrates has phi* attributed phi to O, then (holding constant all the remaining semantic rules of English) "Socrates" would refer to O. So *given* that sentences of that form mean that *Socrates* has phi, it follows "Socrates" refers to Socrates.

Here is the general principle that falls out of this. For "Socrates" to refer to Socrates, it is necessary and sufficient that *Socrates has phi* attribute phi to Socrates. Thus "Socrates" refers to Socrates exactly if *Socrates has phi* attributes phi to Socrates.

Given this, it is tempting to put forth the following generalization:

(A) For any property phi, an expression E refers to an object O exactly if *E has phi* attributes phi to O.

(A) points us in the right direction. But I do not think that, as it stands, it is acceptable. There are two problems with it.

Consider the expression "snores". It is a datum that this expression picks out the property of redness. If you say "that man snores", the word "snores" picks out the property that you are ascribing to the man. So it very

much seems that "snores" refers to a property (that of being a snorer): it is practically a datum that it does so. In any case, a theory of reference cannot presuppose the falsity of that reasonable view.

But (A) doesn't accommodate the fact that "snores" refers to the property of snorer. There are no grammatical sentences of the form *snores has phi*. So (A) doesn't give a general analysis of reference: it gives an analysis that applies only to expressions belonging to a certain grammatical category.

There is another problem with (A). It has to do with the meaning of the word "attributes". That word is too vague; and on one reasonable delineation of it, (A) comes out false. Consider the sentence:

(*) "Today Professor Smith was discussing a great philosopher of antiquity who died of hemlock poisoning."

There is no expression in (*) that *refers* to Socrates. But one could still say that (*) attributes a property to Socrates – that of being discussed by Professor Smith. For any phi, *a great philosopher who died of hemlock poisoning had phi* can be seen as attributing phi to Socrates: but that sentence does not have any component that *refers* to him.

So if (A) is to be turned into an acceptable analysis of reference, two things must be done: it must be broadened to accommodate expressions belonging to diverse grammatical categories; and the concept of "attribution" must be tightened.

Before I give what I believe to be an adequate, and general, analysis of reference, I want to consider one more inadequate analysis. It is worth bringing up because it *seems* very plausible, and because the problems we find in it will motivate the positive analysis I wish to defend.

Dummett once said "the referents of our words are what we talk about". "Socrates" refers to Socrates because we use that expression to make statements about Socrates. If sentences of the form "…Socrates…" didn't say anything about Socrates, then "Socrates" wouldn't refer to Socrates. And given that such sentences are about Socrates, it seems to follows that "Socrates" does refer to Socrates. Given this, it is tempting to say the following: "Socrates" refers to Socrates exactly if, in virtue of having the form "…Socrates…", a sentence is about Socrates. The appropriate generalization is this:

(B) An expression E refers to O exactly if, in virtue of having the form "…E…", a sentence is about O.

(B) has a virtue that (A) lacks: it can accommodate expressions like "snores" and "red" – expressions that *don't* belong to the same grammatical category as "Socrates".

In virtue of having the form "…snores…", a sentence is about the property of snoring. If you say "John snores", your sentence is about the property of snoring: you are saying that John has that property. So (B) accommodates the fact (if it is a fact) that "snores" refers to that property. In virtue of having the form "…red…" a sentence is about the property of redness: if you say "Smith's car is red", your sentence is about that property – you are attributing it to Smith's car. So (B) can accommodate the fact that expressions other than nouns can refer; and, to that extent, (B) seems preferable to (A).

I think that (B) points us in the right direction. But there is a problem with it: the term "about" is too vague. On one delineation of that term, (B) is false. (*) is arguably *about* Socrates. At the same time, there is no expression in (*) that *refers* to him. If a sentence has the form *…a philosopher who died of hemlock poisoning…*, it is, at least arguably, about Socrates. But it is *not* that case that, in virtue of having that form, a sentence contains a component that *refers* to Socrates.

Another example might help. An utterance of:

(#) "A great man who discovered electricity and invented bifocals once said that lawyers shouldn't represent themselves in court".

Arguably, this sentence is about Franklin. But it doesn't contain any expression that *refers* to Franklin. It seems that, in virtue of having the form "…a great man who discovered electricity and invented bifocals…", a sentence is *about* Benjamin Franklin. But such a sentence doesn't necessarily contain any expression that refers to him. So (B) would seem to be false.

Here I should insert a disclaimer. My own view is that, strictly speaking, (#) is *not* about Franklin. Like any other sentence, (#) could be *used* to make a statement about Franklin. I might *use* the sentence "sometimes two and two don't make four" to say that your credit card

company won’t issue you a refund or that Smith won’t be given the promotion he so richly deserves. But *in terms of its literal meaning* "sometimes two and two don’t make four" is not about Smith or your credit card company. I believe that, strictly speaking, (#) is no more about Franklin than "sometimes two and two don’t make four" is about Smith. But I know this will strike many as an extreme position; and a correct analysis of reference certainly shouldn’t prejudge the question whether (#) is "about" Franklin. In any case, the concept of "aboutness" is too vague to have a pivotal role in any analysis of reference. So I will provisionally grant that, for the reasons stated a moment ago, (B) is not acceptable.

Frege had a very reasonable view of reference. I believe that it is false, but we must consider it if we are to produce a better analysis.

Let E be an expression that refers to some object O. In Frege’s view, in virtue of having the form "…E…", a sentence encodes a proposition that has for a constituent, not O itself, but rather a concept that singles out O. Consider the expression "Plato". Frege held (rightly, of course) that this expression refers to Plato. In Frege’s view, it is *not* the case that, in virtue of having the form "…Plato…", a sentence encodes a proposition that has Plato for a constituent. Rather, the proposition has as a constituent some concept that applies uniquely to Plato – some concept like *teacher of Aristotle*. (To simplify discussion, let us suppose that Plato was Aristotle’s only teacher.)

Frege held that definite descriptions are singular terms. So in his view "the teacher of Aristotle was bald" is a singular term that refers to Plato. Thus, in virtue of having the form "…the teacher of Aristotle was bald…", a sentence encodes a proposition that has the concept *teacher of Aristotle* as a constituent. For exactly similar reasons, in virtue of having the form "…the author of the Republic…", a sentence encodes a proposition that has the concept *author of the Republic* as a constituent.

Frege’s analysis of reference seems to be this.

(FR) E refers to O exactly if, in virtue of having the form "…E…", a sentence encodes a proposition that has for a constituent some concept (or "sense") that applies to O and O alone.

Here I will introduce some terminology to avoid verbosity. Let e be an arbitrary expression, and let o be an arbitrary entity. (I am deliberately leaving it open what kind of entity o might be – it could be an individual, a

concept, a property.) Suppose that, in virtue of having the form "…e…", a sentence encodes a proposition that has o as a constituent. In that case, let us say that e "semantically contributes" o. (Sometimes we will just say "contributes".)

So, in Frege's view, E refers to O exactly if E semantically contributes a concept that singles out O. In virtue of having the form "… E…", a sentence encodes a proposition that has as a constituent, not O itself, but some *concept* that applies to O and O alone.

According to the view I will defend, E refers to O exactly if occurrences of E contribute *O* itself. In virtue of having the form "…E…", a sentence encodes a proposition that has O itself as a constituent. But let us focus on Frege's view for the moment.

Frege's view has a great virtue. It explains why:

(*) "the teacher of Aristotle was wise"

has a different cognitive value from:

(**) "the author of the Republic was wise".

In Frege's view, the proposition meant by (*) has as a constituent the concept *teacher of Aristotle*, while the proposition meant by (**) does not have such a constituent. And the proposition meant by (**) has as a constituent the concept *author of the Republic*, whereas the proposition meant by (*) does not have such a constituent. (*) and (**) thus encode different propositions, explaining why they *convey* different propositions – why they have different cognitive values.

But for reasons we've seen, Frege's view is false. Let us focus on the expression "the teacher of Aristotle". What is the concept contributed by that expression? Here a slight detour may be useful.

Consider the sentences:

(i) A teacher of Aristotle was wise
(ii) Three teachers of Aristotle were wise.
(iii) No teachers of Aristotle were wise.

Each makes a statement about the concept: *teacher of Aristotle.* (i)

says that something instantiating that concept was wise. (ii) says that three things instantiating that concept were wise. (iii) says that nothing instantiating that concept was wise.

Clearly, all of "a teacher of Aristotle", "no teachers of Aristotle", and "three teachers of Aristotle" contribute the concept: *teacher of Aristotle*. Perhaps they make other contributions. But that concept is their *common* contribution.

If "the teacher of Aristotle" contributes a concept, it contributes that concept. So, if Frege is right, (*) encodes a proposition that has that concept as one of its constituents; it thus makes a statement about that very concept. In this respect, it is just like (i)-(iii).

Supposing this is right, what is that statement? What does (*) say about the concept: *teacher of Aristotle*?

Russell made it very clear what the answer must be. If the concept *teacher of Aristotle* is uninstantiated, then (*) will be false. On the other hand, if there are multiple (equally contextually salient) instances of that concept, (*) will be false. So that concept must be uniquely instantiated. If that instance should be unwise, then (*) will be false.

Thus, *if* (*) encodes a proposition concerning the concept *teacher of Aristotle*, then (*) is true exactly if that concept is uniquely instantiated and any instance of it is wise.

Even referentiality (those who think that definite descriptions are singular terms) agree that (*) is true exactly under that circumstance.[42] Strawson agrees with Russell that "the king of France is bald" is true just in case there is at least one king of France, at most one of king of France, and any king of France is bald. Russell says that if there is no such object, then the statement is false; Strawson says that if there is no such object, then the statement is abortive. They both agree that it is true exactly if those three conditions are met.

Everybody – referentialists and descriptivists alike -- agrees that *if* "the greatest teacher of Aristotle" contributes a concept or sense, then (*) means:

($*_R$) Exactly one thing was a greatest teacher of Aristotle and any such thing was wise.

But if ($*_R$) gives the meaning of (*), then "the teacher of Aristotle"

doesn’t refer to Plato at all. It is a quantifier. So either it doesn’t refer to anything or it refers to a function. If we say that "the teacher of Aristotle" contributes a concept, then we must embrace Russell’s theory. The essence of that theory is: definite descriptions do not refer to individuals; they are quantifiers.

By exactly similar reasoning, if we say "the author of the *Republic*" contributes a sense, we are saying that it is a quantifier. In that case, it either refers to a function or it doesn’t refer to anything.

If "the author of the *Republic*" and "the teacher of Aristotle" contribute different senses, it is because they *don’t* co-refer. If, as Russell thought, they don’t refer to anything, then they don’t *co*-refer. If they refer to functions, they refer to different functions, as we saw earlier. If those descriptions are quantifiers, they are different quantifiers, and thus don’t co-refer.

A corollary is that replacing a definite description with a co-referring definite description cannot result in a change in meaning. Replacing "the author of the *Republic*" with "the teacher of Aristotle" results in a change of meaning only if those expressions denote *different* entities (different functions). It follows that, if definite descriptions are singular terms, they semantically contribute nothing more and nothing less than their referents.

Let us sum up this leg of the discussion. According to Frege, in virtue of having the form "…the teacher of Aristotle…", a sentence encodes a proposition that has the concept *teacher of Aristotle* as a constituent. We have seen that, if that view is right, then *the teacher of Aristotle has phi* is logically equivalent with: *something x uniquely a teacher of Aristotle, and x has phi.* But in *that* case, "the teacher of Aristotle" becomes a quantifier, as opposed to a singular term.

This point applies (mutatis mutandis) to Frege’s analysis of proper names. For the sake of argument, suppose that, in virtue of having the form "…Plato…", a sentence encodes a proposition that has as a constituent, not Plato himself, but a concept that singles him out.[43] For the sake of argument, suppose that concept is *teacher of Aristotle.* In that case, in virtue of having the form "…Plato…", a sentence encodes a proposition that has that concept as a constituent.

But in that case, by an argument exactly similar to the one given a moment ago, "Plato" becomes a quantifier. Suppose the proposition meant by "Plato was bald" has the concept *teacher of Aristotle* as a constituent. If

nothing falls under that concept, then "Plato was bald" won't be true. If many things fall under it, it won't be true. So exactly one thing must fall under it. If that thing should not be bald, then it won't be true. If that thing is bald, it will be true. So if the proposition meant by "Plato was bald" has the concept *teacher of Aristotle* as a constituent, then that proposition is true exactly if: something x was uniquely a teacher of Aristotle, and x was bald. Thus, if Frege's analysis is right, "Plato" becomes a *quantifier* and not, contrary to what Frege himself held, a singular term.

Let us sum up. Suppose that, in virtue of having the form "…E…", a sentence encodes a proposition that as a constituent a *concept* that singles out O. According to Frege, under this circumstance, E *refers* to O. We have found that, under this circumstance, E does *not* refer to O. E ends up being a *quantifier*, not a singular term. So the Fregean conception of reference is quite untenable. Frege's view – E refers to O iff E contributes a *concept* that singles out O – turns cases of reference into cases of quantification.

Admittedly, our argument is not yet airtight. For, conceivably, one could say that "the teacher of Aristotle" semantically contributes *both* Plato *and* the concept *teacher of Aristotle.* (And one could make a corresponding claim about "Plato".) But this position is a non-starter. If it is right, then "the teacher of Aristotle is wise" encodes a proposition that concerns that concept *and* that concerns Plato. If "the teacher of Aristotle is wise" encodes a proposition about the concept *teacher of Aristotle*, it means: *exactly one thing was a teacher of Aristotle and any such thing was wise.* If "the teacher of Aristotle is wise" encodes a proposition about Plato, it means: *Plato was wise.* If it does *both*, then it must mean: *Plato was wise and exactly one thing was a teacher of Aristotle and any such thing was wise.* But that is obviously not what is meant by "the teacher of Aristotle is wise". To echo what we said a few pages back, the very idea that it has that meaning is too tortured to consider.

There is more to say. If definite descriptions were *both* quantifiers *and* terms that refer to individuals, then "the teacher of Plato" and "the author of the Republic" would not, strictly speaking, be *co*-referring terms. Up to a point, they would co-refer. But past that point, they would either not be referring terms, and thus not *co*-referring terms, or they would refer to different things. In either case, it would be false to say that, on the whole, those descriptions were co-referring terms. Replacing "the teacher of Plato" with "the author of the Republic" would indeed result in a change of

meaning, but only because those two expressions would *not* be co-referring terms. Replacing "the inventor of bifocals" with "the first post-master general", or "Plato" with "the unique thing x such that arithmetic is incomplete and x=Plato", would result in a change of meaning, but only because the pairs in questions didn't really co-refer. CR would not permit the substitution of, say, "Plato" with "the unique thing x such that arithmetic is incomplete and x=Plato". So CR would no longer permit exactly the substitutions that the Slingshot depends on.

The points we've made about definite descriptions generalize without limit. Consider the sentences:

(A) "Rhenates are warm-blooded"
(B) "Chordates are warm-blooded".

We are inclined to say:

"Rhenates" and "chordates" have the same extension but different intensions. They both refer to the same class; but they pick out that class in different ways. Thus (B) has a different meaning from (A). At the same time, (B) is what results when a referring term in (A) is replaced with a co-referring term.

But that is spurious. Suppose that "rhenates" contributes an intension or concept, and not a set. In that case, it contributes the concept: *rhenate* (or *x is a rhenate*). What does (A) say about that concept? It says that anything instantiating it is warm-blooded. So if "rhenates" contributes a concept to (A), then (A) means

(A*) Anything instantiating the concept *rhenate* instantiates the concept *warm-blooded.*

In that case, for reasons discussed earlier, "rhenates" denotes a function F that assigns truth to a class C exactly if all rhenates fall into C.

If "chordates" contributes a concept to (B), then that sentence means:

(B*) Anything instantiating the concept *chordate* instantiates the concept *warm-blooded.*

In that case, by reasoning parallel to that just given, "chordates" denotes a function F* that assigns truth to a class C exactly if all chordates fall into C.

F and F* are different functions. There are epistemically possible worlds where all chordates are warm-blooded but where some rhenates are not warm-blooded. Let w be such a world. In w, (A*) is true and (B*) is false, showing that F and F* are different functions. Thus, if "rhenates" and "chordates" contribute concepts or intensions to (A) and (B), then they refer to *different* things.

Suppose "rhenates" and "chordates" contribute extensions, not intensions. In that case, both (A) and (B) mean:

(C) A_1, A_2…A_n are warm-blooded,

where A_1, A_2…A_n are exactly those individuals that, in actuality, are rhenates/chordates. So if "rhenates" and "chordates" contribute extensions, then (A) and (B) mean exactly the same thing.

So if (B) has a different meaning from (A), that is because (B) is not what results when a referring term in (A) is replaced with a co-referring term. If (B) *is* what results when a referring term in (A) is replaced with a co-referring term, then (A) and (B) mean exactly the same thing. A corollary is that, in (A) and (B), each of "rhenates" and "chordates" semantically contributes nothing more and nothing less than its referent.

Here we must echo a point made a moment ago. Our argument is not yet airtight; for, conceivably, one could say that "rhenates" and "chordates" contribute *both* their intensions *and* their extensions. But this is not tenable. We've seen that if "rhenates" contributes its intension, then (A) encodes (A*), and that if "rhenates" contributes its extension, then (A) encodes (C). So supposing that "rhenates" contributes *both* its extension and its intension, then (A) must presumably have a conjunction for its meaning:

(A**) Anything instantiating the concept *rhenate* instantiates the concept *warm-blooded*; and A_1, A_2…A_n are warm-blooded.

But (A**) is obviously not the meaning of (A). Surely not every sentence of the form "…rhenates…" encodes a conjunction. The idea is simply too tortured to consider. Perhaps any such sentence is *ambiguous* between intensional and extensional readings. But it simply isn't plausible to suppose that each such sentence encodes a *conjunction* of the kind just described. By analogous reasoning, it follows that "chordates" never contributes *both* an intension *and* an extension.

Of course, (A) and (B) don't *have* to be read as universally quantified statements. (A) could be taken to mean: *some rhenates are chordates* or *rhenates have a tendency to be chordates.* Plural nouns, at least in English, are capable of different readings. If you say "people are mean", it is unclear whether you mean *all people are mean* or *many people are mean* or *not all people are mean, but there is a universal tendency for people to be mean.* (A) and (B) are thus ambiguous. (I leave it open whether this is a matter of pragmatics or semantics. An obvious extension of Kripke's (1977) argument would suggest that it is pragmatic.) But for any reading of (A) and (B), replacing "rhenates" with "chordates" results in meaning-change only if "rhenates" and "chordates" denote *different* functions. Really, this is a triviality. A sense or concept just *is* a function: from objects (or worlds) to truth-values. If expressions contribute different concepts, then they denote *different* functions. A corollary is that under no circumstance can meaning be changed by replacing a referring term with a co-referring term. It follows that, if a term refers to something, it semantically contributes that thing and that thing alone.

§ We are now in a position to put forth what I think may be an adequate conception of reference:

(*) An expression E refers to an object O exactly if, in virtue of having the form "…E…", sentence encodes a proposition that has O as a constituent.

(*) is subject to one refinement, corresponding to the points we made earlier about types and tokens. It is sentence-tokens, not sentence-types, that encode propositions. So the right principle is:

(RF) An expression E refers to an object O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition that has O as a constituent.

(RF) accommodates the fact that expressions other than nouns may refer. For example, the expression "snores" is plausibly seen as denoting the property of snoring. (This is not *all* that it does; that is why that expression is not interchangeable with "the property of snoring". But it is hard to deny that *part* of what occurrences of "snores" do is to pick out, and thus refer to, that bodily function.) In virtue of having the form "…snores…", a sentence-token is about that property. A token of "Plato snores" says that Plato has that property, and is therefore about that property (among other things).

Let me provide an intuitive basis for (RF). As Dummett said, if E refers to O, then sentences of the form "…E…" are *about* O. If "Socrates was wise" were not about Socrates, then "Socrates" wouldn't refer to Socrates. So for "Socrates" to refer to Socrates, it is necessary that sentences (or sentence-tokens, rather) of the form "…Socrates…" be about Socrates, i.e. encode propositions that concern him.

But the converse appears not to hold. Arguably, sentences of the form "…a great philosopher who died of hemlock poisoning…" are about Socrates. But they do not, in virtue of having that form, contain any parts that refer to Socrates. So aboutness appears not to be enough for reference.

But a certain, very strict kind of aboutness *is* enough for reference. We know that tokens of "Socrates" refer to Socrates. This is a paradigm of reference. And we know that, in virtue of having the form "…Socrates…", a sentence-token encodes a proposition that has Socrates as a constituent. Any such proposition is about Socrates in a maximally strong sense. (I myself believe that this is the *only* sense in which, strictly speaking, a proposition can be about a thing. But I leave that aside for right now.) So if we confine ourselves to this strong sense of "about", the following seems to hold: E refers to O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition that is about O in the strong sense just mentioned, i.e. that proposition has O as a constituent.

It is easy to find additional support for this view. We know that in *some* cases, E's referring to O involves sentences of the form "…E…" encoding propositions that have O as a constituent. For the sake of argument, suppose that this didn't hold universally. Suppose that, for some values of E

and O, sentences of the form "…E…" encodes propositions that had some *sense* of E, as opposed to E itself, as a constituent.

Let me make this more concrete. Given what Kripke, Strawson, and others have argued, we have good reason to think that, in virtue of having the form "…Socrates…", a sentence-token encodes a proposition that has Socrates himself as a constituent. But suppose that we take a different line with regard to definite descriptions. Suppose we take a Fregean approach. We regard "the main character in most of the Platonic dialogues" as a singular term that refers to Socrates. But we say that, in virtue of having the form "… the main character in most of the Platonic dialogues…", a sentence encodes a proposition that has a *concept* of Socrates, not Socrates himself, as a constituent.

In that case, it seems to me, we are disjunctivizing the concept of reference. As we saw, if the Fregean view is correct, then

(^) "The main character in most of the Platonic dialogues was wise"

encodes a proposition identical with, or logically equivalent to:

(^^) Something x was uniquely a main character in most of the Platonic dialogues, and x was wise.

But we know on independent grounds that

($) "Socrates was wise"

encodes the proposition:

($$) Socrates was wise.

The relation which *Socrates' being wise* has to (^^) is nothing like the relation which his being wise has to ($$). ($$) is true in every world where Socrates is wise, and vice versa. (^^) is false in many worlds where Socrates is wise, and Socrates is not wise in many worlds where (^^) is true. The relationship between Socrates' being wise, on the one hand, and (^^)'s being true, on the other, is highly indirect: *if* Socrates was a unique main

character in most of the Platonic dialogues, *then* Socrates' being wise is necessary and sufficient for the truth of (^^). But the same is true of *anything.* For *any* object x, *if* x was a unique main character in most of the Platonic dialogues, *then* x's being wise is necessary and sufficient for the truth of (^^). So, given Frege's conception of reference, the *semantic* relation between Socrates' being wise and (^^)'s is as remote as that between *Jefferson's* being wise and (^^)'s being true. Socrates' wisdom is relevant to the truth of (^^) only *if* the additional premise is granted that Socrates alone was a main character in most of the dialogues. But that is an entirely *extra*-semantic premise. On Frege's model, you could know everything there was to know about the semantics of (^^) without having the slightest suspicion as to who it was, if anyone, that had that property. So, on Frege's model, the strictly *semantic* relation between Socrates' being wise and (^^)'s being true is minimal: it is no more robust than the strictly semantic relation between Jefferson's being wise and (^^)'s being true.

There is another way to put this. Consider the sentence:

(^^^) "Something x was uniquely a main character in most of the Platonic dialogues, and x was wise."

We *know* that there is nothing in (^^^) which refers to Socrates. In effect, we have just seen that, on Frege's view, the relationship between Socrates' being wise and (^^^)'s being true is just like the relationship between Socrates' being wise and (^)'s being true. So if we accept Frege's model, then we cannot say that there is anything in (^) which *refers* to Socrates. In any case, if we *did* say this, we'd be extending the term "reference" to cover both actual reference, on the one hand, and quantification, on the other. So to avoid this unwholesome disjunctivizing of the term "reference", we must restrict application of that term to the kind of relation that holds between E and O iff, in virtue of having the form "…E…", a sentence-token encodes a proposition that has as a constituent, not a sense of O, but O itself.

What Frege calls "reference" isn't reference at all: it is quantification. When a *sense*, as opposed to the thing picked out by that sense, is what makes it into a proposition, we have quantification, not reference. (In any case, we don't have reference *to the thing* to which the sense applies. We may have reference to a function.) If we were to insist on

describing that relation as one of "reference", we would be rendering the term "reference" ambiguous; we would be undoing the important semantic work that Russell did.

§ Our analysis of reference is embodied in the following dictum:

(RF) An expression E refers to an object O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition that has O as a constituent.

There is an obvious objection to (RF):

Consider the expression "that Socrates taught Plato". This doesn't refer to Plato; it refers to a proposition. But in virtue of having the form "… that Socrates taught Plato…", a sentence-token encodes a proposition that has Plato as a constituent. Surely the proposition encoded in "it is true that Socrates taught Plato" has Plato himself as a constituent. So (RF) is wrong.

No. (RF) is not wrong. (More exactly, it is not wrong in virtue of the fact just identified by the interlocutor – as we'll see in a moment, there are other reasons why it must be qualified.) A token of "that Socrates taught Plato" contains a constituent that refers to Plato; so such a token does refer to Plato *in part.* But it would obviously be false to say that *on the whole* it was an expression that referred to Plato. It is an expression that, on the whole, refers to a proposition about Plato; and it does this by being, in part, an expression that refers to Plato. So (RF) is actually borne out by consideration of that expression.

Still, the objector's point does require us to clarify a point. E is an expression that *on the whole* refers to O exactly if *two* conditions are met. First, in virtue of having the form "…E…" a sentence-token encodes a proposition that is about O (i.e. that has O has a constituent). Second, there is no object O* (distinct from O) such that, in virtue of having the form "… E…", a sentence-token encodes a proposition that is about O*. A token of "that Socrates taught Plato" is not, on the whole, an expression that refers to Plato because there is some object O *other* than Plato such that, in virtue of having the form "…that Socrates taught Plato…", a sentence-token encodes a proposition that is about O*. (Possible values of O* are Socrates and perhaps

the relation of teaching.)

A little while ago we said:

(RF) E refers to O exactly if a sentence-token of the form "…E…" encodes a proposition that has O as a constituent.

A moment later we amended (RF) to deal with the fact that, on the whole, "that Plato taught Socrates" doesn't refer to Plato.

For the next few chapters, I am going to operate on the view that (RF), and the subsequent amendment, are correct.

But I should warn that it is still an approximation. I think that the term "denotation" is actually ambiguous. I believe that (RF) is *one* legitimate disambiguation of it; but it is not the only one.

Acknowledging the ambiguity of the term "denotation" is crucial to answering many difficult questions. Do expressions other than nouns refer? Does the "snores" in "Plato snores" refer to anything? Does it refer to a property or function? Or does it not refer at all? Is the relation that holds between "snores" and (say) the property of snoring entirely different from the relation between tokens of "Plato" and Plato? Are grammatical differences referentially significant? Supose that Smith tokens a "me" and later tokens an "I". It seems a datum, nearly enough, that those tokens co-refer. But they have different grammatical roles. Is there a way of reducing those grammatical differences to *referential* differences? Perhaps, strictly speaking, those tokens *don't* co-refer; perhaps they refer to different functions (each involving Smith). What about connective expressions like "and", "if… then…", "because"? It seems fair to say that "if…then…" refers to the relation of consequence. But is the relation of that expression to the consequence-relation comparable to the relation of tokens of "Plato" to Plato? Or are we dealing with a different relation altogether? Would we be hyper-extending the notion of reference if we said that, indeed, tokens of "if… then…" *do* refer to the consequence relation? And would we be hyper-extending the notion of reference if we say that tokens of "and" *do* refer to some function (e.g. one assigning truth to a pair of propositions or sentences exactly if both of them are true)?

We've seen that forceless sentence-tokens seem to refer to propositions. But what about forced sentence-tokens? Do they refer to anything?

Here is an outline of my answer. On one disambiguation of the term "denotation", grammatical differences are *not* referentially significant. The tokens of "I" and "me" mentioned a moment ago simply co-refer.. They both refer to Smith; neither refers to a function. End of story. On the other disambiguation, grammatical differences *are* referentially significant.

On the one disambiguation, tokens of "snores" refer to *instances* of the property of snoring. On the other, they refer to that property itself.

On the one disambiguation, forced sentence-tokens do not, on the whole, refer to anything (though they obviously have parts that refer). On the other, they can be construed as referring terms – expressions that refer to functions of a very special and exotic kind.

The term "denotation" is given meaning through certain paradigms. We say that "Fred" refers to Fred, and so on. We also say that "red" refers to redness. Our paradigms pull in two subtly different directions.

At this point anything I say on this topic is bound to be obscure; so I will confine myself to a few brief remarks.

Earlier we briefly considered the idea that propositions might be *properties.* The proposition:

(CM) *Brutus murdered Caesar*

is true in a world exactly if that world is a certain way – exactly if the quantum in that world is rippled a certain way. Just as baldness is a property of Socrates, so propositions are properties of worlds. A proposition is true in a world just in case the corresponding property is instantiated there. So propositions are properties and truth is instantiatedness. Just as a person may or may not have the property of baldness, so a world might or might not have the property corresponding to a given proposition. (Of course, some propositions are *necessarily* true. e.g. *squares are square*. But this doesn't indicate that propositions are not properties. After all, some properties are necessarily had by all things. For example, any individual x is such that x either is, or is not square-shaped.)

If this is right, then a proposition is the kind of that thing that can be *instantiated.* This suggests (though it doesn't strictly imply) that the *components* of propositions must also be things that can be instantiated. Here we have a problem. It is generally held that Caesar himself is a constituent of (CM). But Caesar is not the sort of thing that can be instantiated.

This points to a more general problem. A proposition has *structure*; it is essential to a proposition that its components be arranged in a certain way. Otherwise there is no distinction between *John loves Mary* and *Mary loves John.* Given this, if we think of propositions as things that have for their constituents actual bits of the spatiotemporal world – actual people, murders, snorings, and so on – then we end up making propositions be perilously close to the states of affairs that make them true. The proposition *John loves Mary* ends up being a structure that comprises John, loving, and Mary in a certain arrangement; and the order is presumably one whereby John is loving Mary. But in that case, the proposition *John loves Mary* becomes the state of affairs that would make it true; it becomes a *case* of John's loving Mary; so that proposition becomes automatically true. But obviously that proposition could be false. So if we think of individuals as components of propositions, then it becomes impossible to account for the structural properties of propositions without incurring the absurd result that propositions become automatically true.

To sum up: we want propositions to be things that can be *instantiated*; and we want them to be things that don't *have* to be true (except in atypical cases, e.g. *1+1=2*). The idea that *Caesar* himself is a component of (CM) makes it hard to satisfy these requirements.

For these reasons, and others, I will argue that what is a component of (CM) is not Caesar himself, but is rather the *property* of being Caesar. That property can, of course, be instantiated. So (CM) comprises the property of being identical with Caesar, the relation of murdering, and the property of being identical with Brutus. It also comprises various other properties – these extra properties are needed to account for the profound differences between (CM) and *Caesar murdered Brutus.* And (CM) is true exactly if all of its component properties are instantiated.

If (CM) is *true*, that involves some *state of affairs* that has Caesar himself – and not (just) the property of being identical with him – as a component. So, it seems, the thing which has Caesar as a component is a state of affairs, not a proposition.

I *don't* want to say that "Caesar" denotes the *property of* being identical with Caesar. "Caesar" denotes Caesar. That is a datum. If we want to denote that property, we must use an expression like "the property of being identical with Caesar". The expression "Caesar" won't do.

Consider the proposition. "Caesar is snoring". Suppose it is true, i.e.

suppose the associated proposition is true. In that case, Caesar himself – the person, not some concept of that person – is snoring. So there exists a state of affairs of which Caesar is a component. Thus Caesar is a component of any *state of affairs* in virtue of which *Caesar snored* is true. But Caesar is not, I believe, a component of the proposition itself.

Given all of this, here is what I would say. "Caesar" denotes Caesar because, for any sentence-token of the form "…Caesar…", the *truth* of the proposition encoded therein involves a *state of affairs* that has Caesar as a component. I believe that the right analysis of reference is more or less along the following lines:

(RF) Expression E denotes object O exactly if the following condition is satisfied. Consider an arbitrary sentence-token of the form "…E…". That token encodes a proposition P such that any state of affairs S that makes P be true has O as a constituent.

So I *don't* believe that, in the final analysis, Caesar himself is a component of *Caesar snored.* But in what follows, it will expedite discussion if we speak as though he is such a constituent. So for the next few chapters, we will so speak, and at the end of the present volume, we will make the needed corrections.

§ Earlier we tentatively *rejected* the idea that

(1) E refers to O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition that is about O.

We rejected (1) in favor of this one:

(2) E refers to O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition that has O as a constituent.

We rejected (1) because it seems plausible, on the face of it, to say that "a great philosopher who died of hemlock poisoning snored" is about Socrates, even though it contains no expression that refers to him.

I think that (1) and (2) are actually equivalent. So, as I would now like to show, I think that (1) is correct.

We must distinguish two ways that facts can relate to sentences. Consider the sentence

(i) "Sally met a man today".

Suppose that (i) is true.

In that sentence, "a man" does not refer to any man. Russell's[44] argument for this is powerful.[45] Let N be any expression that refers to a man. *Sally met a man but did not meet N* is not self-contradictory. (The asterices are quasi-quotes.) But (i) would be self-contradictory if "a man" referred to some particular man. Thus (i) is not about Smith or Brown or any other man.

(i) is *made true* by the fact that Sally met Smith or Brown or Jones. But it would be folly to say that (i) was actually about Brown or Smith or Jones. You cannot know what a sentence means unless you know what it is about. You can know exactly what (i) means without having any idea who Sally met. So (i) is not about any particular man. When we say that a sentence is about a certain thing, we mean that it encodes a proposition about that thing. So the proposition encoded in (i) is not about Smith or Jones or any other particular man.

Let P be a proposition of the form…*C*…, where C is some concept that applies to Plato and Plato alone. (When I say that P has the form…*C*…, I mean that P has C as a constituent.) Assume that P does not have Plato as a constituent.

If P is true, some fact about Plato will be what makes it true. But this is not enough for P to be about Plato. Some fact about Smith makes (i) true. But (i) is not about Smith. One can understand sentences encoding P without knowing anything about Plato. So, by an argument parallel to the one just given, P is not about Plato. Facts about Plato *make it true.* But it is still not about Plato.

P is true in worlds where Plato doesn't exist. What a proposition affirms doesn't vary from world to world. So what a proposition is about doesn't vary from world to world. Since P is true in worlds where Plato doesn't exist, it follows that P is not about Plato in any world, including this one.

These days, it is said that a sentence can be about a thing in two

quite distinct senses: it can be *de re* about that thing, or it can be *de dicto* about that thing. A sentence is *de re* about O if it encodes a proposition that has O as a constituent. A sentence is *de dicto* about O if it encodes a proposition that has as a constituent some concept C that (uniquely) applies to O. "Plato smokes" is *de re* about Plato, since it encodes the proposition: *Plato smokes.* "Exactly one person wrote several dialogues that are philosophical masterpieces" is *de dicto* about Plato: it is about a *concept* that applies to Plato: *author of several dialogues that are philosophical masterpieces.*

As we've seen, "Exactly one person wrote several dialogues that are philosophical masterpieces" is not about Plato at all. When we say that a sentence is *de dicto* about x or y, that means that it is about some concept that applies to x or y; it does not mean that it is *actually* about x or y.

"A great philosopher who died of hemlock poisoning snored" is *made true* by the fact that Socrates snored. But nothing in that sentence refers to Socrates. The proposition *expressed* by that sentence is *made true* by some fact about Socrates. But, as we just saw, that proposition is not really about Socrates.

I believe that Frege confused these two relations. Consider the sentence:

(*) "Something x was uniquely a teacher of Plato, and x was wise."

That sentence is *made true* by Socrates' being wise; but the proposition encoded in that sentence could be true even if Socrates had never existed, and there is nothing in that sentence which *refers* to Socrates. If Frege's semantics is right, the proposition encoded in

(**) "the teacher of Plato was wise" is identical with, or logically equivalent to, the proposition encoded in (*). So the relation which Socrates' being wise has to the truth of the (*) is identical with the relation which it has to the truth of (**). Where (*) is concerned, that relation doesn't involve the occurrence of anything that refers to Socrates. So, if we accept Frege's semantics, it would be arbitrary to maintain that there was anything in (**) which referred to Socrates. If Frege's semantics is right, then (**) is *made true* by Socrates' being wise; but that sentence is not, strictly speaking, about Socrates, and there is nothing in it which refers to him. So Frege's view that definite descriptions are singular terms is inconsistent with his own

semantics.

Frege says that the relation of "Socrates" to Socrates is just like the relation of "the teacher of Plato" to Socrates. So Frege's (obviously correct) view that "Socrates" refers to Socrates is inconsistent with his own semantics.

We must always keep in mind the distinction between *content* and *truth-maker*. Consider the proposition: *a great philosopher of antiquity died of hemlock poisoning.* Some fact about Socrates *makes it true.* Some such fact is the *truth-maker* of that proposition. But Socrates is not himself a component of that proposition: he is not part of its *content.* You can grasp that proposition in its entirety without having any concept of Socrates.

The second half of this book will be spent arguing that some current orthodoxies in epistemology and the philosophy of mind that are projections of a failure to keep in mind the distinction between content and truth-maker.

§ Is the term "direct reference" a pleonasm?

As we noted, a token of "you", addressed to (say) Benjamin Franklin, semantically contributes Franklin himself, and not a concept that applies to him. This is also true of tokens of "Benjamin Franklin" and possibly of "the inventor of bifocals".

The doctrine that some terms contribute *individuals*, as opposed to concepts applying thereto, is called "direct reference theory".

But really "direct reference theory" should just be called "reference theory". If tokens of "the inventor of bifocals" contribute the concept *inventor of bifocals*, as opposed to Franklin, then they would not refer to Franklin at all; they would be quantifiers, not expressions that referred to Franklin. Reference is always "direct".

The debate between friends and foes of "direct reference theory" is really a debate as to whether there is such a thing as reference; it is not a debate as to the nature of reference.

§ Here it would be appropriate to discuss what Dummett and Tughendat have to say about reference. Both authors claim to be interpreting Frege. But what is relevant here is not whether they have correctly interpreted Frege, but only whether the theories they put forth are correct.

If I am reading him correctly, Dummett (1973: 192-197) unwittingly

puts forth two distinct theories of reference. His first theory appears to be this. When we say that "Plato" refers to Plato, we are presumably saying that sentences of the form "…Plato…" are *ipso facto* about Plato. It would be absurd to say: "Plato" refers to Plato; but if you want to make a statement about Plato, you'd better not use that expression; for sentences of the form "…Plato…" aren't about Plato. It would also be absurd to say that sentences of the form "…Plato…" are ipso facto about Plato, but "Plato" doesn't refer to Plato. It seems, then, that "Plato" refers to Plato just in case sentences of the form "…Plato…" are *ipso facto* about Plato. So we might as well say that for "Plato" to refer to Plato just *is* for it to be the case that sentences having the form "…Plato…" are *ipso facto* about Plato.

Give or take some nuances relating to the word "about", this is pretty much identical with the theory of reference we put forth. But Dummett conflates this plausible theory with a most implausible one.

If the theory put forth a moment ago is right, then "Plato" is really to be defined *contextually*: for "Plato" to refer to Plato is for sentences of the form "…Plato…" to have certain meanings or truth-conditions. (For some reason, Dummett talks about truth-conditions, not meanings. In this context, this difference is innocuous. But, for the record, I think it is preferable to talk about meanings. I think that truth-conditions are too course-grained. The *meaning* of "1+1=2" is different from the meaning of "triangles have three sides". But any circumstance or "condition" sufficient for the truth of the one is sufficient for the truth of the other.) Given what we said a moment ago, it seems fair to say that "Plato" refers to Plato exactly if the occurrence of "Plato" in a sentence S has a certain effect on the truth-conditions of S. This, in turn, makes it seem fair to say that, if an expression E has a *different* effect on the truth-conditions of sentences in which *it* occurs, then E cannot co-refer with "Plato". The analysis of reference that emerges is this:

Let E be an expression that refers to O, and let R be the rules governing how the occurrence of E in a sentence s affects the truth-conditions of S. In that case, for an expression E to refer to O just is for E* to be governed by those very rules – if it isn't, then E* doesn't co-refer with E.*

But there has been a subtle shift from a good theory to a bad one. To

facilitate discussion, let us introduce a bit of terminology. If an expression E has such and such an effect on the truth-conditions of sentences in which it occurs, let us say that E has such and such a "truth-valence".

It would be hard to deny that if sentences of the form "…Plato…" weren't about Plato, then "Plato" wouldn't refer to Plato. It would also be hard to deny that if sentences of the form "…Charlemagne…" were *ipso facto* about Plato, then "Charlemagne" would refer to Plato. So it is hard to deny that an expression refers to Plato exactly if that expression has a certain truth-valence. But it doesn't follow that *every* aspect of a word's truth-valence is relevant to its reference. Words can co-refer but not have entirely coincident truth-valences. What is needed is only that their truth-valences coincide in the relevant respect.

The word "wet" refers to the property of being wet, and so does "the property of being wet". Why does the first refer to that property? Because a sentence of the form "…wet…" *ipso facto* encodes propositions one of whose constituents is the property of wetness. *Mutatis mutandis* "the property of being wet" refers to that property for the same reason. So those words have the same referent because *in the relevant respect* they have the same truth-valence. But *in other respects* they have different truth-valences. Replacing "wet" with "the property of being", or *vice versa*, can strip a sentence of truth-condition and, conversely, can give truth-conditions to a hitherto meaningless heap of words. So it is patently obvious that "wet" and "the property of being wet" have different truth-valences.

This point can be made with more effect with a highly inflected language like Latin. It is pretty much a datum that "Brutus" and "Brutum" and "Brute" co-refer – they all refer to Brutus. But they have different truth-valences. "Brutus est malus" has truth-conditions; "Brutum est malus" does not (since it is poorly formed). So it is easy to over-state the connection between reference and truth-valence.

Suppose that E refers to O. It is true that E refers to O exactly if E has a certain effect on the truth-conditions of sentences in which it occurs – exactly if it has a certain truth-valence. But it doesn't follow that *every* aspect of truth-valence relates to reference. E and E* can have different truth-valences, but the same referent. This will happen if they differ in their *grammatical* properties.

Tughendat (1970) puts forth a theory that is similar to Dummett's second theory. The notion of referring is replaced with the notion of having

the *same* reference. Vicious circularity is avoided because, as we will now see, the notion of having the same reference is defined independently of the notion of reference. E and E* have the *same* reference iff they have the same effects on the truth-conditions of the sentences in which they occur, i.e. if they have the same truth-valence.

If I am not mistaken, the idea behind Tughendat's proposal is as follows. It is a pre-theoretical datum that "Cicero" and "Tully" co-refer. Given that they co-refer, replacing an occurrence of "Cicero" with one of "Tully" will never change truth-value. (For expository reasons, let us leave aside problems relating to non-truth-functional contexts.) So if E and E* co-refer, that is sufficient for their having the same truth-valence.

But the converse also appears to hold. Obviously "Cicero snored" means that Cicero snored. But suppose that "Tully" had a different truth-valence from "Cicero"; suppose, say, that "Tully snored" meant *Plato snored*. In that case, "Cicero" and "Tully" would *not* co-refer: the differences in truth-valence would guarantee that. Thus, we may conclude that "Cicero" and "Tully" co-refer precisely because they have the same truth-valence. In general, E and E* co-refer iff they have the same truth-valence. So the concept of reference is explained, and to some extent eliminated, by the concept of sameness of truth-valence.

In effect, we've already seen the problem with this. The argument just given has two parts. The first is: if E and E* co-refer, then (setting aside problems relating to intensionality) they have the same truth-valence. *That* statement, I believe, is quite right.

The second leg of the argument is: *if* E and E* have different truth-valences, then they don't co-refer. *That* statement, I believe, is wrong. "The property of snoring" and "snores" co-refer but have different truth-valences. Truth-valence is a more fine-grained notion than reference – the former, unlike the latter, has both a referential *and* a merely grammatical component.

I will end with a clarificatory note. I have sometimes identified the "grammatical" differences between expressions with mere differences in "inflection". Strictly speaking, this identification is probably false. Consider "John is tired" and "John was tired". The only difference here has to do with the inflection on the verb. But those sentences are not mere grammatical variants of each other – they are not related to each other as are "I don't have any money" and "I ain't got no money". The difference in inflection between "is" and "was" appears to be *referentially* significant – a different time is

indicated. So inflectional differences aren’t always referentially innocent. This does require that we sharpen our views as to which terms are purely "grammatical" and which are not – we can no longer identify the grammatical with the inflectional. But this does not, I think, impugn our point that the meaning of an expression-token cannot typically be *exhausted* by its referent, since such tokens must also contribute the connective material that distinguishes a proposition from a heap of objects and, therewith, a sentence from a heap of nouns.

Chapter 6 What do sentence-tokens refer to?

§ Let (a)-(c) be sentence-tokens.

(a) "Mozart wrote music or grass is green".
(b) "If Mozart wrote music, then grass is green".
(c) "Possibly, Mozart wrote music".

In none of (a)-(c) does the occurrence of "Mozart wrote music" have assertoric force: it is "forceless". Only the larger host-sentence has such force.[46]

The general principle seems to be this. If s is a sentence-token that occurs as a proper part of some sentence-token S, then s does not have any kind of force.[47]

But, of course, when a token of a sentence that occurs *on its own* – not as a proper part of another sentence-token – it *does* have force. Let

(e) "Mozart wrote music"

be a sentence-token occurring on its own. (e) has assertoric force. It is "forced".

Let us start with things that we *know* for certain to refer. We know that tokens of "Plato", "Socrates", "that Mozart wrote music" refer. Notice

that such tokens can always occur as proper constituents of a sentence-token. The fact that "Plato" can occur as a proper part of a sentence is what makes it possible to use that expression to attribute properties to Plato. You can (grammatically) surround an occurrence of "Plato" with other verbiage and, in so doing, make statements about Plato. This seems inextricably linked with the fact that "Plato refers to Plato.

What is it for tokens of "Plato" to refer to Plato? Here we don't need an *analysis* of the concept of reference: a few platitudes will do.

If a sentence-token has the form "…Plato…", it is *ipso facto* about Plato. A token of "Plato snored" means: *Plato snored.* If it meant that Charlemagne snored, then (holding constant the remaining rules of English), "Plato" would refer to Charlemagne. If a sentence is of the form "…red…" or "…snores…", it is ipso facto about the property of redness or of being a snorer. "Plato snores" is about the property of snoring – it says that Plato has that property. "Smith's car is red" is about the property of redness – it says that Smith's car has that property.

If sentence-tokens of those forms had nothing to do with redness or snoring, then it would be a little unclear how it could be maintained that "red" and "snores" refer to those things. So it seems that, if E refers to x, then sentences of the form "…E…" are *ipso facto* about x.

By the same token, if sentences of the form "…E…" are ipso facto about x, then it becomes a little unclear how we can deny that E, or at least some component thereof (if it is a complex expression), refers to x.

Suppose you learn that, in some dialect D of English, sentences of the form *Argo has phi* mean *Margaret Thatcher has phi*, but that otherwise D is just like your usual version of English. It seems pretty clear that, in D, "Argo" refers to *Margaret Thatcher.* In any case, it isn't clear (to me) what it would really mean to deny that it thus referred. It seems that *if*, in virtue of having the form *Argo has phi* a sentence means: *Margaret Thatcher has phi*, that is enough for "Argo" to refer to Margaret Thatcher.

Let D* be some other dialect of English that is just like your version of English except for the following. In D*, *so and so splores* means: *so and so writes trashy novels.* It seems pretty clear that, in D*, "splores" co-refers with our expression "*writes trashy novels*", and that it therefore refers to the property of writing such novels.

So it seems that if E refers to x, then sentences of the form "…E…" are ipso facto about x. And it also seems that if sentences of the form "…

E…" are ipso facto about x, then E refers to x.

Again, this is probably no *analysis* of the concept of reference.[48] But it is a reasonably plausible statement *about* reference. And it can, I think, serve as a provisional touchstone in our effort to answer the question: "what do sentence-tokens refer to?"

Given this, some case can be made that forceless tokens of "Mozart wrote music" refer to the proposition *that Mozart wrote music.* Consider the sentence-token:

(PS) "Plato snores".

Here we have an expression referring to a person and an expression referring to a bodily function. Now consider the token:

(N) "necessarily, Mozart wrote music".

The occurrence of "Mozart wrote music" is plausibly seen as denoting the proposition *Mozart wrote music,* and the occurrence of "necessarily" is plausibly seen as denoting the property of being necessarily true.

In (PS), we have expressions referring to a person and a bodily function. In (N) we have expressions referring to a proposition and a modal status. In (N), the occurrence of "Mozart wrote music" is obviously there to pick out that proposition. (It isn't there to affirm it—just to pick it out).And the occurrence of "necessarily" is there to pick out a property of propositions. So it's a little unclear how it could be maintained that, in the context of (N), the occurrence of "Mozart snores" is anything other than an expression that picks out a proposition. It is obvious how to apply this line of thought to (c) .

Consider the sentence-token:

(SP) "Plato punched Socrates".

Here we seem to have one expression referring to Plato, one referring to Socrates, and one referring to the two-place relation of punching. That relation can be seen as being attributed to the ordered pair <Plato, Socrates>.

An analogous story can be told of (b). In (b), the token of "Mozart

wrote music" can be seen as denoting a proposition; the token of "somebody wrote music" can be seen as denoting a different proposition; and the token of "if…then…" can be seen as denoting the consequence relation. The consequence relation can be seen as being attributed to the ordered pair of propositions <that Mozart wrote music, that somebody wrote music>.

In any case, it is pretty clear that, in (b), the occurrences of the component sentences are there to indicate certain propositions; they are not there to affirm those propositions (for neither is affirmed) – or, for that matter, to do anything besides pick them out. So it seems reasonable, at least on the face of it, to say that in (b) the component sentence-tokens refer to propositions.

In general, it seems that *some* case can be made that, in at least some contexts, forceless tokens of "Mozart wrote music" refer to the proposition *Mozart wrote music*.[49]

§ Unfortunately, this line of thought has a serious shortcoming. Let us consider an expression that we *know* for certain to refer to the proposition that Mozart snores:

(a) "that Mozart wrote music".

If, in (M), we replace "Mozart snores" with (a), what ends up is nonsense or, at least, is ungrammatical:

(M*) "if that Mozart wrote music, then somebody wrote music".

This suggests that (forceless tokens) of "Mozart wrote music" do *not* co-refer with tokens of "that Mozart wrote music". After all, if you replace a referring term with a co-referring term, the properties of being grammatical and meaningful are preserved. Compare: "Cicero snored" and "Tully snored".[50]

But is this point really so devastating? "I ate breakfast" is grammatical; "me ate breakfast" is not. The latter appears to be what results when a referring term in the former is replaced with a co-referring term. So intersubstitutions of such terms can change grammar into non-grammar. So

given only that (M*) is ungrammatical, we don't have yet have proof that "Mozart wrote music" doesn't co-refer with "that Mozart wrote music".

There is more to say. "Me ate breakfast" very clearly *suggests* a meaning – the very meaning had by "I ate breakfast". And that, surely, is consistent with the supposition that "I" and "me" co-refer. If you replaced the "I" with "Plato" or "the inventor of bifocals", or anything else that didn't co-refer with the "I", the resulting meaning would *not* be that had by (M). Similarly, (M*) clearly suggests a meaning – the very meaning had by (M). And this is plainly consistent with the idea that the expressions involved in the substitution co-refer. So, thus far, we haven't found too much reason to doubt that, in at least some contexts, forceless tokens of "Mozart wrote music" co-refers with "that Mozart wrote music".

§ Consider any sentence-token of which an occurrence is "snow is white" is a proper part. In other words, consider a sentence-token of the form "…snow is white…" (Of course, a token of "snow is white": itself has that form. But let us leave that limiting case aside.) Examples are "grass is green and snow is white", "necessarily, snow is white", and "Bob believes that snow is white". We've seen that the occurrence of "snow is white" in any such token is forceless.

Surely all those sentence-tokens are about the proposition *snow is white.* Granted, they are also about things other than that proposition; they are about snow and whiteness and even (it could conceivably be argued) the relation of property-instantiation. But there can be no denying that, in virtue of containing the sentence "snow is white", a sentence-token is about the proposition *that snow is white*.

This gives us reason to believe that (forceless) tokens of "snow is white" refer to the proposition *that snow is white.*

§ We've seen some reason to believe that forceless sentence-tokens refer to propositions – to the propositions that they mean. What about forceful tokens of sentence? What about (e), for example? Obviously (e) has components that refer. But I think that *as a whole*, (e) does not refer. I would argue that, taken as a whole, nothing having sentential force actually refers to anything.

We've noticed that anything that we *know* for certain to refer can occur as a proper part of a sentence-token -- tokens of "Plato", "Socrates",

"that man over there", "that snow is white". But forceful sentence-tokens cannot thus occur. Consider some token of: "it is false that Mozart wrote music". The occurrence of "Mozart wrote music" does not have force. What has force is the whole token. Consider a token of "possibly, Mozart wrote music". The occurrence of "Mozart wrote music" does not have force: only the whole sentence-token does. The same is of the occurrence of Mozart wrote music" in "either Mozart wrote music or grass is green".

Now let us consider some more difficult cases. Consider a token of: "It is true that Mozart wrote music". Strictly speaking, the occurrence of "Mozart wrote music" here no more has assertoric force than it does in a token of "it is not the case that Mozart wrote music". Consider a token of: "Mozart wrote music and grass is green". Strictly speaking, the occurrence of "Mozart wrote music" here no more has assertoric force than it does in a token of "Mozart wrote music *or* grass is green".[51]

It seems that, whenever "Mozart wrote music" occurs as a proper part of a sentence-token, it is stripped of force. So, it appears, forceful sentence-tokens cannot occur as proper parts of sentence-tokens.[52] This by itself makes it questionable whether forceful sentence-tokens refer. Everything we know to refer can (and typically does) occur as a proper part of a sentence-token. And this fact seems to have deep roots in the concept of reference. Tokens of "Plato" refer to Plato because you can use such tokens to construct sentence-tokens that say something about Plato. If this weren’t possible, it is hard to see in what sense "Plato" could refer to Plato.

Forceless sentences can be, and always are, used to construct sentence-tokens that are about propositions. A token of "Possibly, Mozart wrote music" is about the proposition *Mozart wrote music.* And it is about that proposition precisely because it comprises a (forceless) occurrence of "Mozart wrote music". The general rule seems to be: if s is a forceless sentence-token that expresses the proposition p, and S is a complex (forceful) sentence-token that is built out of s, then S is ipso facto *about* p. So if we say that forceless sentence-tokens refer to propositions, we stay on the right side of everything we know about reference.

But we put ourselves on the wrong side of this knowledge if we say that forceful sentence-tokens refer. First of all, such things can never occur as proper constituents of sentence-tokens. That by itself raises questions. But, as I will argue, there are even more compelling reasons to deny that anything have sentential force refers.

§ There are a few different reasonable answers to the question "what does a forced sentence-token refer to?" Among these answers are: a proposition, a truth-value, a fact or state of affairs. Let's consider these possibilities. Let T be an arbitrary forced token of "Mozart wrote music". Suppose we say that T refers to the corresponding proposition. The first thing to do is compare it to something we *know* to refer to that proposition, e.g. a token of "that Mozart wrote music". Let T* be an arbitrary token of "that Mozart wrote music". There are some obvious differences between T and T*. T* can occur as part of a sentence; T cannot.

Now T certainly has much in common with T*. Both somehow *indicate* the proposition *that Mozart wrote music.* But T does more than this: T *affirms* that proposition – it doesn't merely indicate it. By contrast, T does not affirm (or deny) it; it merely indicates it. It is clear that, by itself, *referring* to a proposition doesn't involve affirming it or denying it. That is why T* can occur in the context of a sentence-token that denies the proposition *Mozart wrote music* (e.g. "it is false that Mozart wrote music") or in the context of a proposition that is neutral on the question whether that proposition is true ("if it is the case that Mozart wrote music, then somebody wrote music") or in the context of a sentence-token that ascribes truth to it ("it is true that Mozart wrote music"). So, clearly, referring to a proposition does not *by itself* constitute affirming it or denying it (or doing anything else with it).

So to the extent that T *does* affirm that proposition, it plainly does something other than just refer to it. I myself think it is plausible, or at least not obviously false, to maintain that *part of* what T does is to refer to it. Obviously T *affirms* that proposition. In order to do that, it must somehow indicate it; and this, it seems to me (though it won't so seem to others), is not so different from referring to it. But it is also plain that T does something *other* than indicating it: it also affirms it. So it would be wrong to say that *on the whole* T is a singular term that refers to a proposition. Of course, T comprises constituents that refer – to Mozart, to music, and probably to the relation of writing. But it is obvious that T is not *on the whole* a singular term that refers to any of these things. It may be that part of what T does is to indicate a certain proposition. But, for the reason just stated, it would be false to say that *on the whole* T is a singular term that refers to that proposition.

§ A similar argument shows that it would be false to say that *on the whole* T is a singular term that refers to a fact or state of affairs.[53] Referring seems to be something you do when existence is taken for granted; affirming is something you do when existence is *not* taken for granted. If you *say* "there is a king of France…", you are, in so doing, *affirming* the existence of something, and thus not *referring* to that thing. (The affirmation of existence can, perhaps, be used as the basis for a later reference to that thing, as in: "there is a king of France. And *that guy* is really well-read." But the affirmation of existence itself is not a case of referring to him.) But if you use an expression that is meant to simply *pick out* or *refer to* that thing, that means that, from your point of view, the question of existence has been settled. So it is clear that in *affirming* the existence of the fact that Mozart wrote music, you are doing something other than *referring* to that fact.

Let T# be a token of "the fact that Mozart wrote music". For the reason just given, it doesn't seem quite right to say that, in using T#, you are *saying* that it is a fact that Mozart wrote music. It seems that you are taking it for granted. The question whether that fact exists has been settled. But when you utter a forced token of "Mozart wrote music", that is plainly because that very question has *not* been settled: in making that utterance, you are in the process of settling it. So to the extent that T *does* affirm the existence of the fact that Mozart wrote music, it does something other than merely refer to that fact. And it is obviously essential to T that it does this extra, non-referring something. What *makes* something be a sentence, as opposed to a nominal (like "that Mozart wrote music") is precisely that it constitutes an explicit affirmation of some fact, as opposed to a presupposition of it. So, once again, to the extent that T affirms the existence of that fact, it does something more, or other than, refer to it. So it would not be correct to say that T was a singular term that referred to a fact.

There is also the suggestive fact that T# can occur as a proper part of a sentence-token, whereas T cannot.

§ What about the idea that T refers to the truth-value True? Let T^ be a token of an expression that we *know* to refer to the truth-value *True* (so far as we know this of any expression): let it be a token of "the truth-value True". First of all, the relationship which T^ bears to that truth-value isn't remotely like the relation which T bears to it. For one thing, T^ can, while T cannot, be a proper part of a sentence-token. If you want to ascribe the property of being

made of peanut-butter to that truth-value, you embed "the truth-value True" in the right kind of sentence-token, to wit: "the truth-value True is made of peanut-butter". If you want to ascribe that property to all propositions of the form *P or not P*, you can do so by embedding that expression in the right kind of sentence-token: "the truth-value True is had by all propositions of the form *P or not P*." But you cannot use tokens (whether forced or forceless) of "Mozart wrote music" in this way.

It is not out of the question to suppose that *part* of what a forced token of "Mozart wrote music" does is to indicate that truth-value. After all, such a token affirms and (it arguably follows) ascribes that truth-value to proposition. So, perhaps, at some level, in some way, a forced token of "Mozart wrote music" refers to the truth-value True. But that is obviously not *all* it does. A token of "the truth-value True" *clearly* refers to that truth-value (if anything does). But *merely* by using a token of "the truth-value True" in a sentence, you have not necessarily *ascribed* that truth-value to anything; perhaps you used that expression for precisely the opposite reason ("*no* proposition of the form *P and not P* has the truth-value True"); or perhaps you are using that expression precisely to express your neutrality as to its applicability ("it is unknown to me whether there are expressions of the form *P and not P* that have the truth-value True"). So it is plain that *merely* by virtue of referring to the truth-value True, you have not affirmed that it applies to some proposition (or that it does not apply). By contrast, merely by virtue of using a forced token of "Mozart wrote music", you have committed to holding that a certain proposition has that truth-value: so you have (a) indicated a certain proposition and (b) affirmed it. So, plainly, by virtue of using a forced token of that sentence, you have done something other than refer to the truth-value True. Perhaps *part* of what you've done is refer to it. But it would clearly be false to say that, on the whole, such a token was a singular term that referred to it. Given any object O, we could by an exactly similar argument show that a forced token of "Mozart wrote music" is not, at least not on the whole, a singular term that refers to O.

Suppose you want to make a statement about the truth-value *true*. What you'll do is use some expression like: "the truth-value *true*." What you *won't* do is use "snow is white" or "grass is green". Those expressions would be absolutely useless; you couldn't possibly use them to talk about truth-values.

Of course, there is an obvious reply:

If I say "grass is green", I am saying something about the truth-value true. I am saying that grass is green has that truth-value. So I can use "grass is green" to make a statement about that truth-value.

For the sake of argument, suppose "snow is white" refers to the truth-value *true*, just like "the truth-value true" (or "ze truth-value true"). In that case, presumably, "snow is white" ought to be able to perform the functions performed by "the truth-value true". But this isn't the case. Consider the sentence:

(lsp) "The last sentence Plato uttered had the truth-value *true*."

If you replace "the truth-value *true*" with "snow is white" or "grass is green", what results is rubbish:

(lsp*) "The last sentence Plato uttered had snow is white".

Of course, one can do some syntax-chopping to make the substitution work; one can say that (pt) is equivalent to (say):

(lsp**) "The last sentence Plato uttered had the same truth-value as *snow is white*."

But clearly the idea that "snow is white" refers to the truth-value *true* is under serious strain. If "snow is white really refers to the truth-value *true*, why is it so hard to it to make it act as though it did? Why is it so hard to make it behave like expressions that we *know*, for certain, to refer to that truth-value?

Of course, one can say:

Semantically 'snow is white' is synonymous with 'the truth-value *true*'. The differences between them that you've just described are pragmatic.

This obviously over-extends the concept of pragmatics. Pragmatics cannot be used to exonerate *every* semantic theory. Suppose I say that "grass

is green" means: *penguins don't fly*. Obviously my theory will conflict with what "grass is green" is usually taken to mean. Suppose I said:

Semantically my theory is right. Admittedly, the facts appear to be discrepant with my theory. But that has to do with pragmatics, not semantics.

Clearly I would be over-using the concept of pragmatics: the problem is that "penguins don't fly" and "snow is white" are *too* remote in meaning. "Snow is white" *seems* no less remote in meaning from "the truth-value *true*". In fact, they are plainly *more* remote in meaning than the first pair; the first pair consists of two indicative sentences; the second consists of a sentence and sub-sentential expression. So it is hard to see how pragmatics could be used to exonerate the view that those two expressions co-refer.

§ Let us consider an objection to what we've said:

Some connectives are truth-functional. "S and S*" is true iff both S is true and S* is true. The truth of the conjunction is a function of the truth-values of its components. So surely the truth-values of S and S* are at least among the things which they semantically contribute in that context. In general, it seems that in some contexts, what a sentence semantically contributes is a truth-value. So you're wrong to say that sentences don't semantically contribute their truth-values.

Here is my reservation about this. Let S be the sentence "George W. Bush smoked a cigarette on April 2, 2005" When it occurs on its own, it is occurring in a truth-functional context, no less than when it is occurring in the scope of a truth-functional operator like "and". But, for reasons we've seen, one can grasp the semantics of that sentence without knowing its truth-value.

Obviously "and" is truth-functional. But this doesn't mean that sentences falling in its scope semantically contribute their truth-values. Under no circumstances are the truth-values of S and S* part of what they *semantically* contribute to "S and S*". If they *did* semantically contribute their truth-values, one couldn't even *understand* (grasp the semantics of) the

conjunction without knowing those truth-values.

"S and S*" is equivalent to: the sequence <S, S*> is such that both its members are true. So "and" can be thought of as a function that assigns truth to a sequence of expressions exactly if both expressions denote true propositions . Let CN be that function. CN assigns truth to some sequences and falsity to others, just as "x is green" assigns truth to Kermit and falsity to Miss Piggy. The truth of "Kermit is green" depends, of course, on whether Kermit has the property of being green. But the semantic contribution of "Kermit" is not the property of being green. The truth of

The sequence <S, S> is such that both its members are true*

depends on whether S and S* have certain truth-values. But they no more semantically contribute those truth-values than "Kermit" semantically contributes the property of being green in "Kermit is green".

There is another way to put this. Consider the sentences:

(i) "John believes that snow is white."
(ii) "it is true that snow is white".

In (ii), it occurs truth-functionally: (ii) depends on whether it denotes a true-proposition. In (i), it occurs non-truth-functionally. But obviously "that snow is white" makes exactly the same semantic contribution in both (i) and (ii). The reason it occurs truth-functionally in (ii) is *not* that it contributes a truth-value; it is that truth is being *ascribed* to it. Whether an expression occurs truth-functionally or not has nothing to do with what it *means*; it has to do with what properties are *ascribed* to the thing it means. We will revisit this later, when we discuss Frege's theory of indirect discourse. The truth of "Kermit is green" depends on whether Kermit is green; the truth of "Kermit is a frog" depends on whether Kermit is true of heart. But "Kermit" has the same meaning in both: it isn't that it means one thing in "green-functional" contexts, and something else in "frog-functional contexts".

§ I should deal with another possible objection to my analysis. I said that if E refers to x, then sentence-tokens of the form "…E…" are ipso facto

about x. Now a forced token of "Mozart wrote music" is about the proposition *Mozart wrote music*: it is an affirmation of that proposition, after all. If this is right, then it seems to follow from my view of reference that it *does* refer to that proposition. But I just argued that it *does not* refer to that proposition (or to anything else).

Any forced token of "Mozart wrote music" is about Mozart, music, and the relation of writing. But surely we don't want to say that "Mozart wrote music" is a singular term that refers to any one of these things. All we must say is that *part* of what it does is to refer to those things, but that *on the whole* it is not a singular term referring to any of them. Similarly, I have no objection to saying that *part* of what such a token does is refer to a proposition. But for the reasons given earlier, I think it is false to say that *on the whole* it is a singular term referring to a proposition (or anything else). Some may balk at this on the following grounds:

Forceless tokens of "Mozart wrote music" look just like forced tokens of that sentence. You've made some case that the former refer to a proposition. But now you say that, while part of what the latter do is refer to a proposition, it is false to say that on the whole they are singular terms referring to propositions. But this line of thought seems to presuppose that forced-tokens of that sentence have a higher degree of semantic complexity than forceless tokens of it. But this is implausible given that, phonetically and acoustically and perhaps otherwise, a forceless token of that sentence is indistinguishable from its forced counterpart.

There is plainly *something* wrong with this line of thought, since there obviously *is* a semantically significant difference between forced and unforced tokens of "Mozart wrote music". The former have assertoric force; the latter do not. So while it is true (perhaps) that orthographically or acoustically, there are no differences[54] between them, there are obviously semantic differences. Not all important semantic distinctions are phonetically realized. Here we have a case in point.

§ We argued that forced-tokens are not *on the whole* expressions that refer to propositions or states of affairs or truth-values. Merely by virtue of referring to a proposition, you have not *affirmed* anything – if you say "that snow is white", you haven't said anything true or false (though you have

denoted something true or false). But *merely* by virtue of saying "snow is white", you have said something true or false. This difference appears *not* to be a merely grammatical one; it is not comparable to the difference between saying "I like to play soccer" and "me like to play soccer" or to the difference between saying "if that snow is white, then something is white" and "if snow is white, then something is white". So it seems false to say that, on the whole, forced sentence-tokens refer to propositions. By an exactly similar argument, we can show that, on the whole, forced-sentence tokens don't refer to states of affairs or truth-values.

In fact, we can show that, for any object O, a forced token of "snow is white" doesn't refer to O. Consider some expression that *uncontroversially* refers to O – for example *O*. Merely by virtue of uttering *O*, you haven't said anything true or false. So utterances of *O* differ in an important respect from utterances of "snow is white". And it seems wrong to see these differences as "merely grammatical".

Here I must insert a caveat. I think that the argument just given has force – and I *tend* to agree with it. But I also think there *may* be some hope for the idea that sentence-tokens do refer. Let t be a forced token of "snow is white". I think there *may* be some object O with the following property. The idea that t refers to O *is* consistent with the idea that t is true or false. But I am not certain about that argument. And I feel that the argument given a moment ago has a fair amount of weight. So I am inclined to think that forced sentence-tokens do not *on the whole* refer to anything. So that will be my official position for the remainder of this work.

Chapter 7 Black's Principle and Church-Black Debate

Black argues that "snow is white" cannot possibly co-refer with "the True" (or "the truth-value *true*" or "the property of being true"). His reason is this. Suppose that "snow is white" and "the True" co-referred. In that case, replacing the one with the other would turn a meaningful sentence into a meaningless one, or *vice versa.* Take the sentence "the author of the Meno was wise". Replace "the author of the Meno" with a co-referring expression, say "the most famous Platonist of all time". The result is meaningful: "the most famous Platonist of all time was wise." But if we take

(A) "If snow is white, then something is white"

and replace "snow is white" with "the True", what results is nonsense, namely:

(B) "If the True, then something is white".

I think that Black is ultimately right. But his argument against Church is not satisfactory as it stands, since his core assumption (replacing a referring term with a co-referring term doesn't affect the degree of meaningfulness) is quite possibly false. Still, a *related* assumption is, I think, defensible and sustains a compelling argument against Church.

Black's principle is this:

(BP) replacing a referring term with a co-referring term cannot turn sense into nonsense.

"If snow is white, then something is white" is meaningful. "If the True, then something is white" is not meaningful. Therefore "the True" and "snow is white" don't co-refer.

There is a contrary principle. I will call it "Russell's principle", since it is found in Russell's *Principles of Mathematics*:

(RP) replacing a referring term with a co-referring term *can* turn sense into nonsense.

For example, "John snores" is meaningful, but "John, to snore" is not.

There are three ways we can go here.

(i) Black's principle is wrong, and tells us nothing about what sentence-tokens refer to.

(ii) Black's principle is wrong, but can be made right through some

kind of alteration.

(iii) Russell’s principle is wrong – "snores" and "to snore" do not co-refer.

Let us start by trying to make a case for (ii).

(A) "John snores".
(B) "John, to snore".

(B) is indeed meaningless. But meaning can be restored through a purely *grammatical* operation. Just appropriately decline the verb.

Let us contrast this pair of sentences with:

(C) "If snow is white, then something is white".
(D) "If the True, then snow is white".

(D) is meaningless, and *no* merely grammatical operation on "the True" will yield meaning. The barrier between (D) and meaning is *not* grammatical.

Here, then, is a weakened version of Black’s principle:

(WBP) Replacing an expression with a co-referring expression may turn sense into nonsense. But sense can always be restored through a mere grammatical change.

A corollary is:

(CWBP) Suppose that replacing e with e* turns sense into nonsense, and that no purely grammatical change can restore sense. In that case, e and e* don’t co-refer.

By (CWBP), "the True" and "snow is white" don’t co-refer.

It should be noted that, on the face of it, (CWBP) appears *not* to rule out that "snow is white" and "that is snow is white" co-refer. Uncontroversially, "that snow is white" refers to the proposition *that snow is white*. Now suppose you replace the "snow is white" in:

(i) "If snow is white, then something is white"

with "that snow is white". What results is:

(ii) "If that snow is white, then something is white".

Intuitively, (ii) seems to be *merely* ungrammatical; and it thus seems to follow that a purely grammatical change – mere deletion of the complementizer – can turn (ii) into a perfectly meaningful sentence. So (CWBP) supports the thesis that sentence-tokens refer to truth-values.

§ Black never says *why* replacing a referring term with a co-referring term ought to preserve meaningfulness. He simply asserts it. Right now I'd like to give what I believe is the motivation behind Black's thesis. This will provide us with a more nuanced conception of the semantic notions that relate to our inquiry.

Here is what I believe to be the rationale behind (BP). First of all, *propositions* are never meaningless; a proposition *is* a meaning. Expressions are meaningful or meaningless. A proposition is the meaning of a sentence – a meaningless sentence is one that does not encode a proposition. Given this, consider the proposition: *that Plato loves Socrates.* This proposition has constituents – Socrates, the relation of loving, and Plato. That sentence can be expressed in a sentence: "Socrates loves Plato". "Socrates" picks out Socrates, "Plato" picks out Plato, and "loves" picks out the relation of loving. Suppose you replace "Socrates" with a co-referring expression. The new expression does the same thing as the old one -- the same constituent is picked out. So the proposition hasn't changed: there has been no change in its constituency. Thus the new sentence and the old sentence bear the same proposition and are equally meaningful.

There is a crucial non-sequitur in this argument. Some background is needed to expose it. A proposition is not a heap of constituents. It has structure; its constituents are ordered in a certain way. (I remain resolutely silent on the nature of this ordering: it is clearly not anything like a mere spatio-temporal ordering.) The ordinal properties of a proposition are *not* reducible to facts about its constituency. The difference between *Plato loves Socrates* and *Socrates loves Plato* is not that they have different constituents.

For the sake of argument, suppose the first proposition consisted of four constituents: Plato, Socrates, the relation of loving, and x – x may be anything you choose. Clearly the set (Plato, Socrates, the relation of loving, x) is no proposition. No matter how many new constituents we add to that set, we won't produce a proposition. So the ordinal properties of a proposition do not reduce to facts about its constituents. It follows that the ordinal differences between *Plato loves Socrates* and *Socrates loves Plato* do not amount to a difference in their constituencies.

This point about propositions has a semantic counterpart. As Frege himself remarked, a sentence is not a heap or sequence of referring terms. "Plato, the property of snoring" is not a sentence. A heap of words is not turned into a sentence by adding a referring term. "Plato, the property of snoring" is not turned into a sentence by adding some third referring expression. Let x be any referring term you wish. *Plato, the property of snoring, x* is not a sentence. (The asterices are meant to be quasi-quotes.) No matter how many referring terms we add to a list of referring terms, the result is never a sentence. So the difference between "Plato snores" and "Plato, the property of snoring" is *not* that the first comprises some referring term not comprised by the second. To turn a heap of referring terms into a sentence, those terms must be ordered or structured in some way; and, as we just saw, that ordering is not accomplished by adding another referring term. (We are not, of course, talking about mere spatio-temporal ordering: given any *spatio-temporal* ordering of "Plato" and "to snore", the result is not a sentence.) This is a consequence of the fact that the ordinal difference between *Plato snores* and *Plato, the property of snoring* is not a difference in *constituency*: adding a referring term to "Plato, to snore" would change constituency, but would not introduce the needed ordinal information.

The difference between "Plato snores" and "Plato, the property of snoring" is not that the one comprises some referring term not comprised by the other. The difference between them lies in the grammatical differences between "snores" and "the property snoring". Those grammatical differences are therefore not referential differences; and those grammatical differences *do* introduce the ordinal content that distinguishes the proposition *Plato snores* from the heap *Plato, the property of snoring.*

It thus follows that (BP) is wrong. Replacing a referring term with a co-referring term *can* easily turn sense into nonsense. For meaning, reference is not enough: we need reference plus grammar. Reference is different from

grammar, and we can destroy grammar without destroying reference. The problem with (BP) is that it embodies the false idea that the difference between a proposition and a heap is that the former comprises some constituent not comprised by the latter. It *also* follows from our discussion that (RP) is right. The grammatical differences between "snores" and "to snore" are not referential. They encode different pieces of ordinal information, but this information does not amount to a difference in reference. Given that (RP) is right, it follows that (iii) is not a viable option.

Finally, it also follows from our discussion that (WBP) *is* correct. Let S be a sentence, and let S* be the corresponding non-sentential heap. (So if S is "Plato loves Socrates", then S* is "Plato, to love, Socrates".) Trivially, no referential changes are needed to convert S* into S. It is also clear that the changes that S* must undergo are confined to facts about inflection and word-order – in other words, they are purely grammatical changes. A corollary is that (CWPB) is correct. It follows, as we saw, that "the True" and "snow is white" *do not* co-refer.

There is an important consequence: *semantic contribution* is not exhausted by *reference.* The "snores" in "Plato snores" doesn't merely refer to a property; it also indicates how that property fits into the proposition in question.

This brings us to what I believe is an important point. It seems pretty clear that at least some expressions have a dual semantic role. "The inventor of bifocals" has both sense and reference. I myself think that sense is a property of the *type* "the inventor of bifocals", whereas reference is a property of its *tokens.* But there is no doubt that *some* kind of semantic two-dimensionalism is needed.

But if I'm not mistaken, our semantics must be *three*-dimensional. There is sense, reference, *and* also the ordinal information we've been discussing.

Recall what we said earlier. The existence of indexicals requires us to take a two-dimensional approach to the semantics of *some* expressions. Considerations of uniformity require us to extend that approach to *all* expressions. So, strictly speaking, it is *tokens* of "Plato" that refer: the type doesn't refer to Plato, and its meaning is a (constant) function that assigns a referent to each of its tokens. This point can, and must, be applied to all words. The meaning of the *type* "and" is a (constant) function F such that, for any *token* of "and", F assigns some specific semantic content to that token.

That semantic content may, in its turn, be a function – one that assigns truth to certain sequences of propositions and falsity to others. The point is that a demand for uniformity demands that we extend our two-dimensionalism to all expressions. In cases where it *seems* not to be indicated, we are dealing with constant, as opposed to variable, functions.

Given this, and given what we said a few pages back, a case can be made for a *three*-dimensional semantics. Consider the expression-type "the relation of loving". Tokens of that type refer to the property of loving. (For the reasons given a moment ago, the type itself does not so refer.) But a token of "the relation of loving" does not *just* refer to that relation. Such a token also contributes *ordinal* information. That is why such a token of "the property of loving" is semantically different from a token of "loves", even though there is no difference in reference.

Consider the sentence-type "Bob loves Sally". The proposition encoded is *that Bob loves Sally.* We've agreed that this proposition has the same *constituency* as *that Sally loves Bob.* The differences between them are purely ordinal. The requisite ordinal information is obviously coded in "Bob loves Sally" – that is why it differs in meaning from "Sally loves Bob" and "Bob, the relation of loving, Sally". In some languages, this information is coded in case-endings. ("Sally" and "Bob", or their translations rather, would be differently inflected.) In English, this information is coded in word-order. Thus, the occurrence of "Sally" in a token of "Bob loves Sally" refers to Sally *and* it also indicates the position that Sally is to have in the corresponding proposition. So that occurrence has a dual semantics. Thus, when we take the type "Sally" into account, we have a tri-partitioning of meaning.

I would suggest that this three-dimensionalism applies universally. Consider the expression "and". Tokens of that type are plausibly seen as denoting a function F that assigns truth to certain sequences of propositions and falsity to others. But that cannot be *all* that such tokens do. Consider the expression "the function F". Tokens of that expression uncontroversially refer to F, but differ grammatically from tokens of "and". The difference lies not in what is denoted, but in the ordinal information encoded. The meaning of the expression-*type* "and" is a function that assigns F, along with the aforementioned ordinal information, to each of its tokens. The meaning of the *type* "the function F" is a function that assigns F, along with different ordinal information, to each of *its* tokens.

Here is the picture that emerges. Expressions have a purely

grammatical component and a non-grammatical component. "Loves" and "to love" are the same in respect of the non-grammatical component, but not in respect of the first. The grammatical component encodes purely information; it doesn't refer to anything. (Remember that the needed ordinal information must not be *referred* to.) The non-grammatical component refers to something. Expression-tokens thus have a dual semantics. Expression-types have a different semantics. Thus we are left with a three-dimensional semantics.

§ We are trying to figure out to what, if anything, sentence-tokens refer. The problem is that, while we have solid intuitions about reference when it comes to *certain* expressions, we have at most weak and indecisive ones when it comes to others. Does "but" refer? What about "and"? Intuition is silent. Intuition is equally silent when it comes to whole sentence-tokens.

But, I believe, an analysis of reference falls out of our discussion of Black's principle. We've made a case that *Plato loves Socrates* has exactly the same *constituents* as *Socrates loves Plato.* In connection with this, we saw some reason to believe that there is no *referential* difference between "loves" and "the relation of loving", even though there is a *semantic* difference. Thus, "loves" refers to the relation of loving because that relation is a *constituent* of the propositions encoded in "Plato loves Socrates", "Socrates loves Plato", and the like. The difference between "loves" and "to love" has nothing to do with what is referred to -- nothing to do with a difference in the corresponding propositional constituents. In both cases we are dealing with the same constituent; what differs is that constituent's *relation* to its fellow constituents. So, to a first approximation, we may analyze reference thus: E refers to O exactly if a sentence-token containing E *ipso facto* encodes a proposition one of whose *constituents* is O.

This last statement has to be refined. If the expression "that snow is white" occurs in a sentence S, S *ipso facto* encodes a proposition one of whose constituents is snow. But *the* referent of "that snow is white" is a proposition, not snow. This is easy to correct. There is a unique x such that, if S comprises "that snow is white", the proposition meant by S *ipso facto* comprises x and such that, for any y such that S *ipso facto* includes y, y is a constituent of x. x, of course, is the proposition *that snow is white.* So E refers to O exactly if two conditions are met (we will refer to this analysis of reference as R1).

R1

(i) A sentence S comprising E *ipso facto* encodes a proposition one of whose constituents is O.

(ii) Let x be any entity such that, if S comprises E, then S *ipso facto* encodes a proposition one of whose constituents is x. In that case, x is either identical with O or is a constituent of O.

Here there is a Fregean objection:

"Plato" refers to Plato, and presumably "Plato snores" means Plato snores. But the proposition Plato snores doesn't have Plato as a constituent. Rather, it has a sense that applies (uniquely) to Plato.

If this were right, then "Plato" would be a quantifier, not an expression that refers to Plato. Remember what we said earlier. Consider the definite description "the murderer of Smith". That description has a sense: *murderer of Smith.* In so far as that sense makes it into the proposition meant by "the man who murdered Smith is rich", the definite description is functioning attributively. In so far as the definite description is functioning attributively, that sentence means: *somebody uniquely murdered Smith and any such person is rich.* So in so far as it is functioning attributively, "the man who murdered Smith" is a *quantifier,* not an expression that refers to Plato. Now suppose that "the man who murdered Smith" is functioning *referentially* in "the man who murdered Smith is rich". In that case, the concept *man who murdered Smith* serves *not* as a constituent of the proposition, but merely as a way of *picking out* some such constituent. Of course, the constituent in question is Fred; and the proposition in question is simply *Fred is rich.* So in so far as the definite description is functioning referentially, Fred himself is a constituent of the proposition meant by "Fred is right." Thus in so far as the concept *man who murdered Smith* is part of what is meant by a token of "the man who murdered Smith is rich", the definite description is a quantifier, not a singular term. In so far as the definite description is not functioning as a quantifier, that sentence-token means simply: *Fred is rich,* Fred himself being a constituent of its meaning.

An exactly similar argument applies to "Plato". For the sake of argument, suppose with Frege that "Plato" has a sense or concept – say, *inventor of the theory of Forms.* In that case, "Plato" is synonymous with "the inventor of the theory of Forms". This means we can map everything we just said about "the murderer of Fred" onto "Plato". In so far as the concept *inventor of the theory of Forms* makes it into the proposition meant by "Plato snores", then "Plato" is functioning attributively and, therefore, as a quantifier, not a singular term. In so far as "Plato" is functioning referentially, Plato himself is a constituent of what is meant by "Plato snores".

It has been said that, when "Cicero" occurs in an "oblique" context, what is *ipso facto* a constituent of the sentence in question is not Cicero, but is rather some sense. (This is meant to explain the obvious differences in cognitive value and the alleged differences in truth-value between "Fred believes that Cicero is Tully" and "Fred believes that Cicero is Cicero".) But when that is maintained, it must *also* be maintained (as it always is) that, in such a context, "Cicero" doesn't refer to Cicero – it refers to a sense. So if that really is what "Cicero" does in oblique contexts, then our principle is actually confirmed.

Reference is not an entirely theoretical notion – the layman routinely says things like "what are you referring to?" But R1 has a rather theoretical flavor, and probably doesn't immediately resonate with our pre-theoretical conception of reference. Right now I'd like to delineate the roots that R1 has in that conception. This will lead us to what appears, at first, to be an alternative and perhaps superior (because less tendentious) analysis of reference. But we will see that this alternative is either unacceptable or that it collapses into R1.

Echoing Frege, Dummett says "the referents of our words are what we talk about". "Plato" refers to Plato because, if a sentence contains "Plato", it is ipso facto *about* Plato. In general, E refers to O if a sentence containing E is ipso facto *about* O.

This is subject to a qualification exactly similar to the one stated earlier. If a sentence contains "that snow is white", it is ipso facto *about* snow; but "that snow is white" refers to a proposition, not to snow. Here is how we deal with this (let us refer to this conception of reference as R2):

R2: E refers to O exactly if two conditions are met:

(i) if S comprises E, then S is *ipso facto* about O.

(ii) Let x be any entity such that, if S comprises E, then S is *ipso facto* about x. In that case, x is either identical with O or is a constituent of O.

I think that R2 corresponds pretty well to our pre-theoretical notion of reference. "Plato" refers to Plato because, if a sentence-token has the form "…Plato…", it is *about Plato.* If sentences having that form were *ipso facto* about Charlemagne, then "Plato" would presumably refer to Charlemagne. This point works in the other direction: if sentences of the form "… Charlemagne…" were *ipso facto* about Plato, then "Charlemagne" would refer to Plato. Another illustration: Sentence-tokens having the form "… red…" are *ipso facto* about the property of redness. If such sentences were ipso facto about Plato or about the property of green, then "red" would refer to that person or that property. So it seems E refers to O exactly if sentence-tokens having the form "…E…" are *ipso facto* about O. Thus R2 apparently corresponds to our intuitions about reference fairly well.

R2 might seem superior to R1 in certain respects. R1 is committed to the controversial doctrine known as direct-reference theory: when "Plato" occurs in S, Plato himself is a constituent of the proposition meant by S. By contrast, R2 seems neutral on this. For "Plato" to refer to Plato, it isn't necessary that Plato be a constituent of the proposition meant by "Plato was bald. The proposition need only be *about* Plato: it is left open *how* this "aboutness" is to be accomplished. Perhaps Plato himself is a constituent, perhaps not – perhaps only a sense that applies to Plato is such a constituent.

But I think the advantage that R2 has over R1 is illusory, and that R1 is preferable. Suppose that Fred kissed Sally and that no one else did. In that case, the sentence "some man kissed Sally" is true. Nothing in that sentence *refers* to Fred.[55] At the same time, it is not unreasonable to say that that sentence is *about* Fred. The expression "some man" is such that, in virtue of its occurrence there, that sentence is "about" Fred. But "some man" doesn't refer to Fred. So R2 seems too permissive.

The counter-response is to say this. "Fred kissed Sally" is "about" Fred in a far stricter sense than "some man kissed Sally". The latter isn't really about Fred at all. It is *made true* by some fact about Fred. But it is not really about him, at least not in a strict sense. There are possible worlds where "some man kissed Sally" is true, but where it was Jerry, not Fred, who

kissed Sally. By contrast, there are no worlds where "Fred kissed Sally" is true but where Fred did not kiss Sally. So "Fred kissed Sally" is *about* Fred in a far stricter sense than "some man kissed Sally". So, after all, it *is* true to say that E refers to E exactly if sentence-tokens of the form "…E…" are ipso facto about O, with the qualification that we are using the word "about" in the *strict* sense.

But what is that strict sense? It seems to me that Fred is a *constituent* of the proposition *Fred kissed Sally* but not *some man kissed Sally*. Indeed this point is almost a truism. These days it would be said that "some man kissed Sally" is *de dicto* about Fred, whereas "Fred kissed Sally" is *de re* about Fred. This means that the first is about a proposition or *concept* (a dictum) that applies to Fred, whereas the latter is about Fred himself: Fred himself, and not merely a concept applying thereto, is a constituent of the proposition affirmed. This is why Fred exists in any possible world where the first is true, but not so with the second. So when we make it clear what the sense of "about" is, we recur to R1.

The main reservation people will have about R1 is the idea that, if an expression referring to Plato occurs in a sentence-token, then the proposition thereby meant actually has Plato himself as a constituent. But many will insist that it is not Plato himself, but rather a *concept* of Plato, that occurs in that proposition. This is because many are uncomfortable with the idea that a physical *object* could be "trapped" (as Kaplan[56] put it) in a proposition, and would prefer to have a sense there in his stead.[57] For reasons already given, I would insist that Plato *himself* must be a constituent – otherwise "Plato" becomes a *quantifier*.

Church on Black

We've seen that Black's principle is probably not quite right as it stands, but that, with a few changes, it can be turned into something correct.

Alonzo Church had an entirely different reaction to Black's point. Church simply rejects it wholesale. In a review of Black's article "Frege on Functions", Church writes:

A more serious error, in the reviewer's opinion, is the fallacious attempt to refute Frege's view, that sentences are designations of truth-

values, by reference to the grammar of English language. It is pointed out that if the sentence "Three is a prime" is a designation of the True (Black uses a capital letter in translating Frege's *das Wahre*), then the expressions "Three is a prime" and "the True" ought to be interchangeable in non-oblique contexts – as indeed Frege himself maintained in analogous cases. Thus from "If three is prime then three has no factors" we get "If the True then three has no factors." To the latter expression Black objects calling it nonsense, and by saying that it "has no more use than" the expression "if seven then three has no factors." In the absence of supporting reasons for such objections, it must be supposed that Black is rejecting the expression in question on the ground that it violates the rules of English grammar. But surely the right question to ask here is not what, by existing custom, are the rules of English, but rather what it is desirable to take as the rules of a formalized language. In a suitable formalized language the analogue of "if the True then three has no factors" does have a use, namely as a designation of the False. And indeed it is not unusual, in formulations of propositional calculus, and of formalized languages containing propositional calculus, to introduce primitive constants denoting one or both of the two truth-values, and to allow substitution of such constants (as well as of longer sentences) for the propositional variables.[58]

Church seems to be saying that the meaninglessness of "if the True, then three is prime" is a mere fact about English grammar. But that is not so. The meaninglessness of that expression reflects deep logical facts. Those facts are, in their turn, *reflected* in facts about grammaticality. But we are not dealing with *mere* ungrammaticality.

Consider a case of a sentence-token that can plausibly be regarded as *merely* ungrammatical:

(1) "John no like play soccer".

(2) is ill-formed. But it is clear enough what it means or, at any rate, what it is supposed to mean: *John doesn't like to play soccer.*

But

(2) "if the True, then three is prime"

is in a very different category. The connective "if…then…" expresses a relation between truth-evaluable entities (sentence-tokens or propositions); it expresses the relation of consequence or entailment. The True is not itself true or false; neither is "the True"; therefore those things entail nothing and nothing entails them. Only propositions (or, if you are a nominalist, sentences-tokens) stand in entailment relations. The True is not such a thing. The meaninglessness of (2) obviously reflects that fact. That is not a peculiarity of grammar. It is a deep fact about the concept of entailment.

Church talks about the "analogue" of (2) in formal languages. If by "analogue" Church means "translation", then Church's point is spurious. Translations preserve meaning – or the lack thereof. Any correct translation of (2) will be as meaningless as (2). So Black's point will apply to any translation no less than to (2).

Of course, we could use (2) to mean:

(3) *Given any true proposition P, P entails that three has no factors*.

Perhaps when Church talks about "analogues" of (2), he is referring to sentences that mean (3). But when (2) expresses a proposition like (3), "the True" doesn't function as an expression that refers to the property of truth. It functions rather as a quantifier. (Thus, depending on one's views on quantifiers, it is functioning either as something that doesn't refer at all or as something that refers to a second-level function.)

The difference between "snow is white" and "the True" isn't remotely like the difference between "I" and "me". In the one case, we are dealing with something that is plausibly seen as a mere grammatical difference. In other cases, we are dealing with a much deeper difference: a difference that is reflected in, but not confined to, a grammatical difference.

There is a subtler point to make here. Arguably, grammatical differences *are* semantic differences. There are grammatical differences between "red" and "John". You can say "John smokes", but not "red smokes". ("The property of redness smokes" is false, perhaps even absurd, but perfectly grammatical. But "red smokes" is simply ungrammatical.) But these grammatical differences presumably reflect semantic differences. It could even be maintained (in fact, I will maintain it later) that these semantic differences reduce to differences in what the expressions in question *pick out.*

The idea would be that the one picks out an individual, while the other picks out a function.[59]

To take another example: There are profound grammatical differences between "or" and "Smith". But those differences obviously reflect semantic differences. And a case *can* be made (I will try to make it) that those semantic differences reduce to differences in what those expressions pick out. ("Or" picks out a higher-order function, "Smith" picks out an individual.)

It is by no means clear that grammatical facts can ever be dismissed as *merely* grammatical – as semantically innocent. On the face of it, they seem to embody semantic facts; and grammatical categories seem to be *semantic* categories. When you say that "John" and "some man" are "noun phrases", you are presumably making a statement about the semantic roles of such expressions. By this reasoning, when you say that "I" and "me" are grammatically different, you are saying something about their semantic roles.

The temptation is to say that "I" and "me", and "eager" and "eagerly", differ *merely* in ways that are grammatical. But that is, I believe, quite false. I would suggest that there are subtle semantic differences between (say) "I" and "me". The function of a token of "I" is not *merely* to contribute an individual. That is merely a good first approximation. It would be very strange if it turned out that deeply entrenched grammatical facts – like the difference between "I" and "me" – were semantically innocent.

Obviously when semanticists produce semantic models, they have to abstract from subtleties – just as physicists must initially idealize away from wind-resistance and the like. But we must not confuse the model with the thing modeled. In some of our models of English, there is no place for a difference between "I" and "me". But that doesn't mean that there *is* no semantic difference. On the face of it, there almost certainly is. In any case, what I am saying is that we cannot assume that grammatical facts are of no semantic consequence. It is quite possible that grammatical differences *do* reflect semantic differences. In that case, Black's point prevails *even if* the difference between "snow is white" and "the True" is "merely" grammatical.

§ There is another point to make about Church's response to Black. On the face of it, Church isn't really responding to Black: he is simply changing the subject. Church seems to be saying:

When we are designing a formal language, it is sometimes

convenient to suppose that sentences of that language denote truth-values. That is what is important. It isn't important what sentences of natural language refer to. So let's forget about natural language, and focus on the needs internal to the architectonic of a new and improved semantics of our own making.

Semanticists – most notably Frege and Church – have indeed claimed it "fruitful" to suppose that sentences of formal languages "denote" or "refer to" truth-values. They have said that various desiderata relating to uniformity, compositionality, and the like are naturally satisfied by *stipulating* that sentences of such languages "denote" truth-values. This thinking is generally thought unexceptionable. Who can argue with a stipulation? If Church *stipulates* that sentences of a formal language of his own making are to denote truth-values, then how can we possibly refute him? How can a convention be refuted? Of course, we *could* take issue with his assertion that the needs internal to his system legitimate that particular stipulation: we could say that, given the formal properties of that language, some other stipulation would be preferable. But who would want to argue with Church about the logical structure of his own language? Who, indeed, would want to go toe to toe with Church in the arena of formal logic? And who would dare confront him regarding the logic of *his own language*?

I believe that this combination of reasoning and fear has led philosophers to *concede* to heavyweights like Church that, in formal languages, it is legitimate to stipulate that sentences refer to truth-values – indeed, it is legitimate to *stipulate* that they refer to anything we want them to.

But not all stipulations are legitimate; some are illogical. An order cannot be true or false, but it can embody a falsity (e.g. "build me a round square"). The same is true of a stipulation. We've seen what it is for an expression to denote something. E denotes O exactly if, in virtue of having the form "…E…", a sentence-token says something about (attributes some property) to O. "Plato" refers to Plato because *Plato has phi* means that Plato has phi. Now imagine an artificial language L with the following properties. The sentences of L have the same meanings as their English counterparts; so "snow is white" means *snow is* white, and so on. But in L, by our stipulation, "snow is white" denotes the truth value of its meaning. That

stipulation is incoherent. In virtue of having the form "…snow is white…" a sentence (whether of L or of English) encodes a proposition that has the proposition *that snow is white* for a constituent; such a sentence does *not*, at least not in virtue of having that form, have *the truth-value True* for a constituent. That by itself means that the L-expression "snow is white" *does* refer to a proposition and does not refer to a truth-value.

An indicative sentence is something that is true or false. (To simplify discussion, let us leave aside non-indicative sentences.) The truth or falsity of a sentence turns on the truth or falsity of the corresponding proposition. Let E be an arbitrary sentence in an arbitrary language L – L may be artificial or natural – and let P be the proposition meant by E. Let S be a sentence of L of which E is a proper part. In virtue of having the form "…E…" S will say something about P – P will be a constituent of the proposition meant by S. It is *not* the case that, in virtue of having that form, a truth-value is a constituent of that proposition. It is not up to us to *stipulate* what sentences refer to. It is not up to us to *stipulate* what E denotes in that sort of context. Once we have assigned a proposition to E, the matter is out of our hands. When E occurs as a proper part of another sentence, it will denote a proposition, not a truth.

And it is equally out of our hands what E denotes when it occurs on its own. When a sentence-occurs on its own, it has *force*. Something that has force cannot occur as a proper part of another sentence. And if something cannot thus occur, then it lacks the defining features of anything that we *know* to refer. For this reason, and others, we earlier saw it necessary to *deny* that whole sentences, occurring on their own, are referring terms. We cannot stipulate this fact about reference out of existence. For E to refer to O just *is* for it to be the case that, if a sentence-token has E as a proper part, then that token encodes a proposition that has O as a constituent. If we are designing a formal language, then any expression s in that language that qualifies as a sentence will satisfy the following requirements: First, when occurring on its own, s is true or false, and thus has some proposition p for its meaning (in some sense of the word "meaning"). Second, when s occurs as a proper part of a sentence S, S encodes a proposition P that has p as a constituent.

So once we endow s with enough semantics to make it be a sentence, in any reasonable sense of the word, then we are endowing it with *too* much semantics to refer to a truth-value. Given this, suppose that after giving s the properties needed to qualify as a sentence, we "stipulate" that s denotes a truth-value. In that case, our stipulation runs contrary to the

semantics that we have given s. It is like stipulating that x is to be a square, and then stipulating that x is to be a triangle. So Church's reply is faulty even relative to his restricted purpose of designing formal languages, and is also faulty relative to Black's purpose of giving the semantics of English.

Chapter 8 Are all terms referring terms?

Suppose that, in virtue of having the form "…E…", a sentence-token encodes a proposition that has O as a constituent. In that case, let us say that occurrences of E "semantically contribute" (or, simply, "contribute"). I have repeatedly said (though not always in these exact words) that

(SC) E refers to O iff E semantically contributes O.

But (SC) is in apparent contradiction with a very popular theory But I doubt our arguments for SC have addressed everyone's misgivings about it. The main misgiving, I think, is this:

Maybe you're right to say that tokens of "Plato" refer to Plato exactly if Plato is semantically contributed by such tokens. So in some cases meaning collapses into reference. But SC says that meaning always collapses into reference, and this is plainly false. Consider the word "and". This word has meaning. But surely it doesn't refer to anything. It has meaning by virtue of having some kind of syntactic or formal function. It does not have meaning by virtue of picking out some object.

You have a Neanderthal conception of meaning: to mean is to denote. This is exactly the position that Wittgenstein refuted in both the Tractatus and in his later works. In the Tractatus, he refuted it in connection with connective terms like "and" and "or". In his later work, he refuted it in connection with lexical items like "Nixon" and "Socrates".

Semantic content is not *always* identical with reference. Expression-types, I have argued, do not refer to anything; but they have semantic

contents: functions of some kind.

But with that qualification, I do have exactly the conception described. Where expression-tokens are concerned, to mean is to denote. When we say that some expressions have a purely "formal" or "syntactical" function, we are really saying that they denote special kinds of things: second-order functions, to be precise.

Consider the word "and". The standard view about it is this:

You don't define "and" by pointing to some object that it denotes. You define it contextually. You say what it means by saying what is meant by whole sentences containing it.

The idea seems to be that if "and" referred to something, you could define it non-contextually; you could simply indicate what it denoted. Since it must be defined contextually, it doesn't refer to anything.

This view is sheer folly. For "Smith" to pick out Smith is *precisely* for "Smith" to admit of a certain kind of contextual definition. It is precisely for "…Smith…" to mean:…*Smith*…

Of course, "Smith" can be defined ostensively, whereas "and" cannot. But ostensive definition is a form of contextual definition. When you say:

(*) *That* person is named "Smith"

that is just a condensed way of giving a contextual definition. You are saying:

(**) Consider *that* person over there. Let O be that entity. "… Smith…" means:…*O*…

This can be shown by analogues of arguments already given. Suppose you point to some object O and you say:

That person [pointing at O] is named "Smith". But if you want to make a statement about him, you cannot use the word "Smith" to do so. "Smith smokes" doesn't mean that he [pointing at O] smokes; it means that

some other person smokes. In general "…Smith…" doesn't mean that he [pointing at O] has:…x…

That would be absurd. You started out by giving an ostensive definition. But you then stripped that definition of any force it initially had.[60] So if you point to O and say "*that* is Smith", a necessary condition for your definition to have any force is that "…Smith…" mean:….*O*…

That is also a sufficient condition. Suppose you said:

Sentences of the form "…Smith…" mean that that person [pointing to O] has:…x…But "Smith" does not name that person.

That would be absurd, for now familiar reasons.

When you give an ostensive definition, you are really giving a condensed contextual definition.

Of course, pointing cannot be involved in a definition of "and" or "or". But that has nothing to do with semantics; that has to do with metaphysics and epistemology. We cannot point to abstract objects or to objects existing in the future. But we can refer to such things.

Perhaps we must define "and" contextually. But this is of no semantic consequence. "Smith", no less than "and", is always defined contextually.

Strictly speaking, we can no more point to the meaning of "Smith" than we can point to the meaning of "and". When we say that Smith is the meaning of "Smith", we are really making a statement about a class, an infinitely large class, of expressions. Smith's being referred to by "Smith" – Smith's being the meaning of "Smith" – is really identical with the fact that sentences of the form "…Smith..." mean:…*Smith*.…We can more point to *that* fact than we can point to the meaning of "and".

The other side of the coin is that any contextual definition can be turned into a non-contextual or denotative definition. Frege rightly said that "some man" does not denote some man. He concluded that it must be defined contextually:

(i) "…some man…" is true exactly if: *for some x, x is a man and*.… *x*…

This is what people typically say about "and" and case-markers, and other so-called non-denoting expressions. This is taken to imply that such expressions do not denote anything. But it is easy to convert any contextual definition into a denotative definition. For example, (i) is equivalent to:

(ii) For any concept C, *C(some man)* is true exactly if for some x, x is a man and C(x).

(ii) is equivalent to:

(iii) for any concept C, "some man" assigns truth to C exactly if for some x, x is a man and Cx.

(iii) is equivalent to:

(iv) "some man" denotes a function that assigns truth to a concept C exactly if, for some x, x is a man and Cx.

What we just said about "some man" can be done with *any* expression. Any contextual definition can be reduced to a denotative definition. If an expression can be defined contextually, it denotes something.

In effect, we've already seen this. Connectives like "after", "and", and "because" denote functions. The same thing is true of tense- and case-markers.

Granted, there is a difference between expressions like "Plato" and expressions like "and". But the difference is not that the former denote, while the latter do not. The difference lies in *what* they denote. The former denote individuals. The latter denote functions.

Actually, we will see some reason to believe that the difference is even *more* tenuous than this suggests. Arguably, tokens of "Plato" *do* denote functions, no less than "and". The difference between tokens of "Plato" and tokens of "and" lies in what kind of functions they denote.

The Augustinian conception of language, ridiculed by Wittgenstein, turns out to be quite right.

Since reference reduces to semantic contribution, it follows that sentence-tokens refer to what they semantically contribute; they refer to what they mean. In any case, if one is to escape this conclusion, one must show

that SC is wrong. But this does not appear to be possible.

Is it a truism that meaning is compositional?

Many (most famously Frege) have maintained that the meaning of a sentence is a function of the meanings of its parts. Linguistic meaning is "compositional". This is known as the "thesis of compositionality" or, simply, "compositionality".

Compositionality has been accepted by most, but not all. To my knowledge, it has always been regarded as a substantive thesis. I think that compositionality is correct. But I also think it is a truism, in the same category as "we *can* travel in time: at the rate of one second per second".

We have argued that, for (tokens of) "Socrates" to refer to Socrates just *is* for sentences of the form "…Socrates…" to have a certain meaning. If they *don't* have that meaning, then "Socrates" didn't refer to Socrates to begin with. There is no way that a sentence-token could not be a function of the meanings of its parts. An expression means such and such exactly if sentences containing it have certain meanings. So whatever meanings those sentences have, that fixes what the meaning of that expression is. Under no circumstance can we meaningfully, let alone truly, say that E means O but that "…E…" does not mean:…*O*…Compositionality is not a thesis about subsentential meaning; it merely identifies what subsentential meaning is.

Chapter 9 The confusions underlying the sense-reference distinction

According to Frege[61], Mill[62], Church[63], and others "the phi" is a referring term that *also* semantically contributes its sense.

We've already seen why their position is wrong. But that position is not to be deplored. For it satisfies several important desiderata. First, it accounts for the fact that

(i) 'the inventor of bifocals snored"

has a very different cognitive value from

(ii) "the first post-master general snored".

Further, it accounts for that fact *without* taking the radically counter-intuitive view that definite descriptions do not refer to individuals.

Russell's theory accounts for the difference in cognitive value between (i) and (ii). But it pays a heavy price: it denies that "the number two" refers to the number two. On Russell's theory, "the number two" refers to nothing or it refers to a function of some kind.

But there is a *third* desideratum that must be satisfied by any correct semantics for definite descriptions. That semantics must not conflict with SC. The Frege-Church position conflicts with SC. That is why it is wrong.

Russell's position is consistent with SC, but it doesn't satisfy our second desideratum.

There is a way to satisfy all these desiderata. We need only make it clear what the difference is between those who hold that definite descriptions are referring terms and those who hold that they are quantifiers. Let us describe the latter as "Russellians" and the former as "referentialists".

The difference between referentialists and Russellians is subtler than is usually thought. The only difference is that, in the semantic rule for "the phi", the referentialist puts the "means that" operator a few inches to the right of where the Russellian puts it.

According to Russell, the semantic rule for "the inventor of bifocals" – the rule that assigns it meaning – is this:

(RUS) Sentences of the form "…the inventor of bifocals…" mean: Exactly one individual O invented bifocals and…O…

According to the referentialist, the semantic rule for "the inventor of bifocals" is this:

(REF) If there is exactly one individual O such that O uniquely invented bifocals, then sentences of the form

"…the inventor of bifocals…" mean: …O…

And if there is no unique individual who invented bifocals, then

"…the inventor of bifocals…" doesn't encode any proposition.

In general, according to the Russellian, the semantic rule for "the phi" is: *"…the phi…" means: Exactly one individual O has phi and…O….*

According to the referentialist, the semantic rule for "the phi" is: *If there is a unique individual O having phi, then "….the phi…" means: …O…*
And if there is no individual uniquely having phi, then "…the phi…" doesn't have any proposition for its literal meaning.

(RUS) and (REF) are very similar. The only difference lies in where the "means that" operator is placed. (REF) gives it *narrow* scope. (RUS) gives it wide scope. Apart from that, (RUS) and (REF) are in near perfect agreement.

But this difference in scope is important. An immediate consequence of it is that, for advocates of (REF), "…the phi…" is *de re* about the unique individual (if any) having phi, whereas for advocates of (RUS), "…the phi…" is only *de dicto* about that individual.

Both (REF) and (RUS) agree that "the phi" is semantically associated with descriptive information. In fact, both (REF) and (RUS) agree that "…the phi…" will be true if and only if there is exactly one object O having phi and…O…

Advocates of (REF) couldn't agree more that "the phi" is semantically associated with descriptive content; that, as a matter of semantics, in order for "…the phi…" to be true, it is necessary and sufficient that there be exactly one object O having phi and that O have…x…

The difference between (REF) and (RUS) is that, for (REF), this descriptive content has a reference-*fixing*, not a meaning-*giving* role.

Here, of course, I am alluding to Kripke's (1972) important distinction. If I say "Gödel was the person who proved the incompleteness of arithmetic", I am telling you who "Gödel" refers to; I am not saying that the *meaning* of "Gödel" includes anything having to do with mathematical proofs. More accurately, I am saying (the asterices are meant to be quasi-quotes):

(*) Somebody O uniquely proved that arithmetic is incomplete and, for any predicate *…x…*, *…Gödel…* means:…*O*…

I am *not* saying:

(**) For any predicate *…x…*, *…Gödel…* means: *somebody O uniquely proved that arithmetic is incomplete and…O…*

(REF) says: "the inventor of bifocals" is a referring term. As a matter of semantics, tokens of it pick out that individual who uniquely invented bifocals, and it picks out nothing if there was no such individual. So for (REF) the descriptive content associated with "the inventor of bifocals" serves to pick out *to whom* "the inventor of bifocals snores" attributes the property of snoring. But that descriptive content does not *itself* figure in the proposition encoded in that sentence.

(RUS) says: that descriptive content *does* figure in that proposition. That sentence says of that descriptive content that it is uniquely instantiated and that anything instantiating it snores.

But notice that, if (REF) is right, then anyone who has the foggiest notion of the semantics of "the inventor of bifocals" knows that "the inventor of bifocals" will certainly not be *true* if there is no inventor of bifocals[64], and also knows that it will not be true if there are multiple (equally contextually salient) inventors of bifocals. So if (REF) is right, then when you say to me[65]:

(IBS) "The inventor of bifocals snores"

I know that, if your utterance is to be (literally) true, the following must be the case:

(*) Somebody uniquely invented bifocals and any such individual snores.

If (REF) is right, then (IBS) will *communicate*, though it won't semantically encode, that a certain description is uniquely instantiated, and any instance of it snores. So a proposition concerning bifocal-inventing will certainly be *communicated.*

So the supposition that (REF) is correct is compatible with several important facts. First, (IBS) is true exactly if somebody uniquely invented bifocals. Second, (IBS) communicates (though it doesn't semantically encode) the proposition: *somebody uniquely invented bifocals, and any such person snored.* Third, the association between the expression "the inventor of bifocals" and the corresponding descriptive information is *semantic*: as a matter of semantics, "the inventor of bifocals" refers to an individual exactly if he uniquely invented bifocals.

Let us now return to Frege and Mill. Frege and Mill want to satisfy three desiderata:

First, they want to do justice to the fact (ISB) *communicates* something about bifocal-inventing.

Second, they want to do justice to the fact that there is a *semantic*, not a purely circumstantial, connection between the description *inventor of bifocals* and the expression "the inventor of bifocal".

Third, they want to say that "the inventor of bifocals" is a term that refers to the inventor of bifocals – not a second-order function.

There is a very natural way to accommodate all of these desiderata. It is to take the position that the semantics of "the inventor of bifocals" is given by (REF).

The other side of the coin is that if, like Frege and Mill, you believe that definite descriptions are referring terms, then it becomes extremely difficult to accommodate (i) and (ii) without taking a position more or less like (REF).

An illustration

"The richest man in Holland" seems to be a term that refers to an individual. But there is plainly a sense in which one can understand "the richest man in Holland sleeps in a sensory deprivation tank" without having

any idea who the richest man in Holland is.

Some (Frege, Church) have reacted by saying that you can know what a referring term means without knowing what it refers to. But that is absurd.

Others (Russell) have reacted by saying that "the richest man in Holland" doesn't refer to the richest man in Holland.

Neither reaction is necessary. According to the referentialist, "the phi" is defined thus: *If there is a unique O such that O has phi, then "…the phi…" means:…O…If there is no such object, then "…the phi…" encodes no proposition.*

So anyone who knows the semantic rule for "the phi" – anyone speaks English --knows that "the phi has psi" will be true exactly if there is a unique phi and any phi has psi. So if you speak English, then you automatically know that

(RMS) "The richest man in Holland sleeps in a sensory deprivation tank"

will be true iff exactly one person is a richest man in Holland and any such person sleeps in a sensory-deprivation tank. It immediately follows that RMS communicates the proposition:

(RMP) exactly one person is a richest man in Holland and any such person sleeps in a sensory-deprivation tank.

(RMP) is exactly the proposition that, according to Russell, is the literal meaning of (RMS). Russell's main reason for saying that (RMP) gives the meaning of (RMS) is that the latter communicates the former. But this fact favors the referentialist's position as much as it favors Russell's. The referentialist's position, no less than Russell's, has the consequence that (RMS) communicates (RMP).

The referentialist does not say:

(i) "the richest man in Holland" means: *Jan van Freejling.*

The referentialist says:

(ii) *If* there is some object O that is uniquely a richest man in Holland, then

"…the richest man in Holland…" means:…*O*.…And if there is no such object, then "…the inventor of bifocals…" has no proposition for its literal meaning.

If (i) were the referentialist's position, then it would be impossible to explain the difference in cognitive value between "the inventor of bifocals snored" and "the first post-master general snored". But (ii), not (i), gives the referentialist's position. Opponents of the referentialist view seem to think that (ii), not (i), is a fair statement of that position.

Unlike the Russellian position, the referentialist's view is *obviously* consistent with a lot of other facts about definite descriptions. For example, if you want to assert that exactly one person is a richest man in Holland, you would actually *avoid* using the term "the inventor of bifocals"; and this, obviously, is *prima facie* inconsistent with Russell's theory.

Maybe the Russellian can solve this problem. But it is a problem for his theory.

The referentialist position doesn't even have this problem. According to (REF), (RMS) doesn't assert the existence of anything. *If* a certain thing exists, *then* (RMS) makes a statement about that thing. If that thing doesn't exist, no statement is made. This corresponds to our intuition that, if there is no richest man in Holland, RMS is abortive, not true or false; and also with our intuition that you cannot use RMS to *assert* the existence of a unique richest man in Holland.

According to Russell's theory,

(RMQ) "Does the richest man in Holland sleep in a sensory deprivation tank?"

is ambiguous between:

(RMQ_W) Is it the case that somebody is uniquely a richest man in

Holland, and that any such person sleeps in a deprivation tank?

and

(RMQ_N) Somebody x is uniquely a richest man in Holland: Does x sleep in a sensory deprivation tank?

But when you token RMQ, you are not asking whether there is a richest man in Holland, nor are you asserting as much. In any case, this is what our intuition tells us.

The Russellian may be able to finesse this problem. But he does *have* it. (Actually, the Russellian cannot finesse this problem, as we'll see.)

The referentialist doesn't even have that problem. For him the rule for RMQ is this:

If there is exactly one richest man O in Holland, then "Does the richest man in Holland sleep in a sensory deprivation tank?" means: Does O sleep in a sensory deprivation tank? *And if there is no such object, then that question is abortive; it fails to ask anything.*

This fits very closely with our intuitions. RMQ *does* seem abortive if there is no richest man in Holland. RMQ seems *not* to assert whether a certain individual exists; nor does it seem to ask whether there is such an individual.

The reason Russell's theory is given credence is exactly that it explains the cognitive difference between, for example, "the inventor of bifocals snored" and "the first post-master general snored". But the referentialist position explains this difference equally well; and it also fits with the other data we've described. And, of course, the referentialist's view, unlike Russell's, harmonizes with our intuition that "the inventor of bifocals" refers to the inventor of bifocals. So it's not clear what evidence favors Russell's view.[66]

Here is another way to look at it. By our stipulation, "ze phi" refers to the unique phi, if there is such a thing, and to nothing if there isn't. No evidence could provide any support for the view that "ze phi" is a quantifier.

It is an analytic proposition that "ze phi" is *not* a quantifier; and nothing can constitute support for the negation of such a proposition. But the evidence that supposedly favors Russell's theory equally favors the view that zefinite descriptions are quantifiers. "Ze inventor of bifocals snored" and "ze first post-master general snored" differ in cognitive value in exactly the same way as "the inventor of bifocals snored" and "the first post-master general snored". That difference in cognitive value provides no support for the view that zefinite descriptions are quantifiers. So the corresponding fact about definite descriptions provides no rational support for Russell's theory.

Here is yet another way to look at it. Consider a community of people who speak a language we will call English*. English*-speakers speak and write exactly as we do, with this one possible difference. Wherever we use a definite description, an English*-speaker uses a zefinite description. *But* English*-zefinite descriptions are homonymous with English definite descriptions: they sound the same and are written the same. So the English*-translation of "ze inventor of bifocals" is "the inventor of bifocals". It seems to me that any evidence favoring the view that we speak English also favors the view that we speak English*, and thus that definite descriptions are zefinite descriptions – that the referentialist is right.

Semantic versus conceptual analysis

It is a matter of controversy whether Russell's theory is correct. But it is generally agreed that it is a reasonable theory. Obviously there are *prima facie* problems for it. For if Russell's theory is correct, then

(i) "is the man in the corner a professor?"

is ambiguous between:

(ii) is it the case that there is exactly one (contextually salient) corner x, and exactly one (contextually salient) man y in x, and that y is a professor?

(iii) There is exactly one (contextually salient) corner x: is it the case that there is exactly one (contextually salient) man y in x and that y is a professor?

(iv) There is exactly one (contextually salient) corner x and exactly one (contextually salient) man y in x: is it the case that y is a professor?

So, on Russell's theory, when one utters (i) one is either asserting that there are a certain number of contextually salient men and corners, or one is *asking* about the number of contextually salient men and corners. But (i) does not *register* with anyone as having any of these meanings. Russell's theory requires us to say that (i) has these meanings; but there is no direct evidential support that (i) has any of these meanings. For (i) is absolutely never used to ask any of (ii)-(iv).

Further, the referentialist's theory, as we saw earlier, simply bypasses this problem. For the referentialist "the man in the corner" is a referring term, and the semantic rule for (i) is this: *If* somebody x uniquely satisfies the description *man in the corner*, then (i) means:

(v) is x a professor?

If there is no such man – if the only inhabitant is a woman dressed as a man, or is a hallucinatory man, or there are several equally salient men in that corner -- then (v) is abortive.

Our intuition strongly favors the view that it is (v), and not any of (ii)-(iv), that gives the meaning of (i).

Russellians acknowledge how counter-intuitive it is to suppose that any of (ii)-(iv) are what is meant by (i). But they say, over and over, that this is not *really* a problem for Russell's theory. For Russell's theory concerns *semantics.* Semantics is not pragmatics. What is literally meant is often not what is communicated. So we shouldn't put too much stock in these deviations between literal and communicated meaning.

Of course, there is often a huge gulf between literal and communicated meaning. But pragmatics is not a blank check. The deviations happen in accordance with certain laws. It is very easy to show that consideration of pragmatics proves the *falsity*, not the truth, of Russell's theory.

To do this, we must distinguish between *semantic* analysis and *conceptual* analysis. The reason Russell's theory has been given credence is, precisely, that this distinction has not been clearly made, at least not in

connection with that theory.

To give a *semantic* analysis of a sentence is to say what it means. A semantic analysis of:

(i) "is the man in the corner a professor?"

says what its literal meaning is. A *conceptual* analysis of a sentence does something else entirely. If P is the proposition meant by some sentence S, a *conceptual* analysis of S gives non-trivially necessary and sufficient conditions for the truth of P. Consider the sentence:

(vi) "Alpha is a circle",

where Alpha is some diagram. The proposition meant by (vi) is obviously:

(vii) Alpha is a circle.

A conceptual analysis of (vi) gives some proposition P* that is distinct from, but equivalent with, (vii):

(viii) Alpha is a closed planar figure of uniform curvature

(ix) Alpha is the class of all points equidistant from a given point in a plain.

There are, indeed, infinitely many different analyses of (vii). (When I say "different analyses", I mean *non-trivially* different.)

It is very seldom said explicitly whether Russell is doing conceptual or semantic analysis. Often, when Russell's theory is attacked, the Russellian counter-attack is to say: Russell is doing *conceptual* analysis; he isn't doing linguistics. When Frege analyzed numbers as sets of equipollent sets, he wasn't doing linguistics. He was doing conceptual analysis. That is what Russell is doing with "the phi".

But there is no doubt that Russell is doing *semantic* analysis; and once this is granted, his theory collapses.

For the sake of argument, suppose Russell is doing *conceptual*, not

semantic analysis, of:

(x) "the inventor of bifocals snored".

That means he is trying to give necessary and sufficient conditions for the truth of the proposition *encoded* in that sentence. What is that proposition? There are only two options. Consider *any* noun-phrase you please: "Socrates", "the man in the corner", "some man", "two penguins". Given *any* noun-phrase, that expression is, at the level of literal meaning, either a referring term or a quantifier. So "the inventor of bifocals" is *either* a singular term – one that refers to Benjamin Franklin -- or a quantifier. If it is a singular term, then (x) semantically encodes the proposition:

(xi) *Franklin snored.*

(xi) does not entail, and is not entailed by,

(xii) *somebody x uniquely invented bifocals and any such person snored.*

So *if* "the inventor of bifocals" is a singular term, then Russell's theory, considered as a *conceptual* analysis of (x), is an unmitigated failure. The proposition that Russell's theory associates with (x) is no more equivalent with the proposition encoded in (x) than is *bananas are a good source of potassium.*

So *if* Russell's theory is to be a correct *conceptual* analysis of (x), then "the inventor of bifocals" cannot be a singular term. It must be a quantifier: for any noun that is not a singular term is a quantifier. Which quantifier is it? For reasons discussed earlier, there is no doubt that *if* "the inventor of bifocals" is a quantifier, it is exactly the quantifier Russell has in mind. So *if* "the inventor of bifocals" is not a term that refers to Franklin, then it is, at the level of literal meaning, the kind of quantifier Russell has in mind. In that case, the literal meaning of (x) is (xii). So *if* Russell's theory is to be a correct *conceptual* analysis, then it is necessary that the *literal meaning* of "…the phi…" be: *exactly one x has phi and…x…*

So Russell's theory is a correct *conceptual* analysis only if it is a correct *semantic* analysis – only if it gives literal meaning.

So when things go badly for Russell's theory, we cannot say: that theory is a piece of conceptual, not semantic, analysis. Russell's theory must be dealt with *as* a piece of semantic analysis.

So Russell's theory supposedly gives the *literal meaning* of (i) and (x). What does this entail? It means that if (i) has a *different* meaning from (ii)-(iv), then Russell's theory is wrong; it means that if (x) has a different literal meaning from (xii), then Russell's theory is wrong. So if Russell's theory is right, then (i) is *synonymous* with (ii)-(iv), and (x) is *synonymous* with (xii).

To say that two sentences are synonymous is to make a fairly strong claim. To say that two sentences are synonymous is to say that they are but different phonetic, or acoustical, variants of each other.

(xiii) John is an enemy

and

(xix) John is a foe

are synonymous exactly because they are but different phonetic representations of the very same thing.

Now that we are clear about what Russell's theory is really saying, it is easy to evaluate it and, indeed, to show that it is false.

Kripke on descriptions

Kripke (1977) makes exactly the opposite claim. Kripke acknowledges that there is some *prima facie* evidence suggesting that Russell's analysis is wrong. For example, *prima facie* it does seem that, when you say, "the man over there smokes", you are not saying: *somebody* or other is a unique man over there and any such man smokes. Rather, you are saying something of the form: so and so smokes. Kripke acknowledges these facts are *apparent* counterexamples to Russell's theory. But he thinks that none of these apparent counter-examples are *actual* counter-examples.

His argument is this. Let English_R be a language that is exactly like English except for this one possible difference. In English_R, sentences of the form "…the phi…" are to be analyzed exactly as Russell says their English

homonyms are to be analyzed. So in English$_R$, by our stipulation, "…the phi…" means: *exactly one thing x has phi and…x…*

According to Kripke, the position that English$_R$ definite descriptions are quantifiers would be subject to exactly the same *apparent* counter-examples as the thesis that English definite descriptions are quantifiers. In English$_R$, says Kripke, "…the phi…" would communicate the same thing as its English homonym; and there would be no discernible difference between English and English$_R$ in respect of when it would be appropriate to say "…the phi…" as opposed to "there is some x such that x uniquely has phi and…x…" We know that English$_R$ definite descriptions *are* quantifiers. Any *apparent* counter-examples to that thesis are *merely* apparent. So the above *apparent* counter-examples to Russell's theory don't show that Russell is wrong. After all, if *by our stipulation*, we were speaking a language of which Russell's theory were true, the same apparent counter-examples would arise. Therefore those apparent counter-examples are innocuous. A corollary is that those apparent counter-examples to no degree show that we are *not* speaking English$_R$.

On independent grounds, Kripke happens to think that we are *not* speaking English$_R$. But he thinks that the just-mentioned *apparent* problems for Russell's theory are not reasons to reject Russell's theory. Russell's theory can deal with them by citing facts about implicature and pragmatics.

I would suggest that those apparent problems *are* actual problems for Russell's theory. Kripke hasn't taken his thought-experiment far enough.

Let English$_S$ be a language just like English with, at most, this one possible difference. In English$_S$, sentences of the form "…the phi…" mean exactly what Strawson says they mean. So, in English$_S$, the semantic rule for "the king of France is bald" is this: *if* O is a unique king of France, then "the king of France" is bald means: *O is bald.* If there is no unique king of France, then "the king of France" is abortive.

Let us compare English$_S$ and English$_R$. In English$_R$

(i) "The king of France is bald"

is a mere notational variant of:

(ii) "There is exactly one king of France and any such entity is bald".

The difference between (i) and (ii) would be merely phonetic; it would be like the difference, in actual English, between a Southerner's pronunciation of "Harvard" and a Northerner's pronunciation of it. For obvious reasons, in English_R, if somebody said:

(ii) "There is exactly one king of France, and any such entity is bald",

it would be quite appropriate, to respond by saying:

(iii) "Strictly speaking, that is false; for there is no king of France."

Therefore, in English_R, it would be quite appropriate to utter (iii) in response to (i). In English_R, (i) would differ from (ii) only in the sense in which a Southerner's pronunciation of (i) would differ from a Northerner's. And *that* difference would not be enough to make (iii) be any more, or any less, appropriate as a response to (i) than (ii). So, in English_R, no facts about pragmatics or implicature would be needed to explain why that response sounded strange: for that response *wouldn't* sound strange.

But in English – I mean *actual* English -- if somebody said (iii) in response to (i), it would sound strange. The strangeness would have to be explained. But if Russell's theory is right, then (ii) and (i) are mere phonetic variants; they are comparable to differences in accent or pronunciation. Mere phonetic differences are not going to account for the inappropriateness that characterizes saying (iii) in response to (i).

Granted, Russell's theory recognizes many different paraphrases of (i) "the king of France". It can be paraphrased thus:

(iv) "For some x, given any y, y is a king of France exactly if x is a king of France, and anything that is a king of France is also wise."

There are as many Russellian paraphrases of (i) as there are different ways of expressing the universal and existential quantifiers. And even more paraphrases can be generated depending on whether we allow ourselves use

restricted quantifiers ("for any king of France x") or whether we use unrestricted quantifiers ("for any x, if x is a king of France"). We've focused on only one paraphrase of (i); and this might seem to vitiate our argument.

But given *any* Russellian paraphrase of (i), we can construct an exact analogue of the argument we just gave.

In English, it is unquestionably appropriate to say (iii) in response to (ii), but not appropriate to say (ii) in response to (i). According to Russell, (iii) and (ii) are but notational or phonetic variants; they differ only in the sense in which a Northerner's pronunciation of a sentence differs from a Southerner's. And that difference is not enough to generate the difference in appropriateness between saying (iii) in response to (i) and saying (ii) in response to (i).

So, contrary to what Kripke says, the prima facie problems that face Russell's theory would *not* recrudesce if we were speaking English_R. If we were speaking English_R, then "…the phi…" would be a mere notational or phonetic variant of some existence claim; and mere phonetic differences are not enough to generate the marked difference in appropriateness, in some contexts, between saying "…the phi…" and "for some x, given any y, y has phi iff y=x and…x…". If we were speaking English_R, we would know it; the behavior of definite descriptions would make it amply clear.

If Russell's theory is right, then

(v) "Does the maid know we're firing her?"

would be ambiguous between

(vi) "Somebody x is uniquely our maid: does x know we're firing her?"

and

(vii) "Is it the case that we have a unique maid x and that x knows we are firing her?"

According to Russell's theory, (v) is a mere phonetico-acoustical variant of either (vi) or (vii), depending on how it is disambiguated. But,

quite obviously, there will be contexts where (v) would be appropriate where either of (vi) or (vii) would be dramatically inappropriate, and *vice versa.* The Russellian must say that mere phonetic differences – differences that don't have the slightest semantic dimension – are responsible for these gross deviations. But that is simply not the case. Of course, there are various different Russellian paraphrases of (vi) and (vii). But an analogue of this point goes through for any of them.

Let us now talk about English$_S$. But remember that, in English$_S$, the meaning of (i) – or, strictly speaking, of the homonym of (i) – is given by the rule: *if* there is a unique king of France x, *then* (i) means: *x is bald*; and if there is no such entity, then (i) is neither true nor false. So, in English$_S$, (i) is neither true nor false; it *presupposes* existence, and if that presupposition is false, then *no* statement is made. So in English$_S$, it would be very strange to say (iii) in response to (i). Of course, in *actual* English, if you said (iii) in response to (i), your response would sound very strange: strange in exactly the same way as an English$_S$ utterance of (iii) made in response to (i). I conclude that if we take Kripke's thought-experiment to its logical conclusion, we find that the obvious *prima facie* problems for Russell's theory are not mere pragmatic epiphenomena; for they do *not* recrudesce when we consider languages of which, by our stipulation, Russell's theory is correct. And they *do* recrudesce when we consider languages of which, by our stipulation, *Strawson's* theory is correct. So Kripke's thought experiment shows, if anything, that *Strawson's* theory, not Russell's, is consistent with the facts about how definite descriptions behave.

Indirect discourse

Even though the inventor of bifocals is identical with the first post-master general, the proposition communicated by the sentence

(1) "John believes the inventor of bifocals snored"

may certainly differ from that communicated by:

(2) "John believes the first post-master general snored."

In epistemic contexts, replacing "the phi" with "the psi" tends *not* to

preserve the identity, or the truth-value, of what is communicated.

This fact is certainly part of what motivated Frege to say that, in epistemic contexts, "the phi" contributes its sense, not its referent. It is obviously part of what motivated Russell to say that "the phi" is a quantifier, not a referring term.

My own view is that (1) and (2) have exactly the same literal meaning. I have not the slightest doubt that replacing "the phi" with the "the psi" often dramatically affects the identity, and thus the truth-value, of the proposition that is *communicated*.

But this fact to no degree warrants views like those of Russell and Frege. There are, at least, two reasons for this. First, these days, we have very good reason to believe that "Cicero" and "Tully" are semantically nothing but labels, and that, in terms of literal meaning and therefore truth-value, "… Cicero…" coincides exactly with "…Tully…".

But there can be no doubt that the proposition communicated by

(3) "Smith believes that Cicero was a great orator"

can be different from the proposition communicated by:

(4) "Smith believes that Cicero was a great orator"

For this reason, of course, the truth-values of the things communicated by (3) and (4) can differ.

We don't want to say that "Cicero" and "Tully" are quantifiers, or that, in (3) and (4), they contribute their "senses". These moves are non-starters, given what Kripke (1972, 1978) said.

So the very thing that we said a moment ago about definite descriptions in epistemic contexts applies equally to names in such contexts. The very facts which warrant the view that definite descriptions are quantifiers warrant the view that proper names are as well. But we know, with nearly complete certainty, that proper names are *not* quantifiers. So the facts in question – the facts relating to apparent substitution-failures of definite descriptions in epistemic contexts -- do not support the view that definite descriptions are quantifiers.

Also, we know from Kaplan (1989) that indexicals are probably

directly referential. But substituting indexicals with co-referring indexicals results in just the kind of apparent divergence in meaning and truth-value that we just considered.

A token of

(i) "His pants are on fire

encodes the same proposition as a token of

(ii) "Your pants are on fire"

if the indexicals ("he", "you") co-refer, as they well may. But obviously (i) and (ii) communicate very different propositions.

A corollary is that a token of

(iii) "I know that his pants are on fire"

may communicate a proposition having a different truth-value from

(iv) "I know that your pants are on fire"

So exact analogues of the arguments that "prove" that definite descriptions are quantifiers also "prove" that indexicals are quantifiers. But indexicals are not quantifiers. So the arguments in question don't prove anything about definite descriptions.

There is another reason *not* to take those apparent substitution failures as having any semantic significance. By our stipulation "ze phi" is a term that refers to the unique phi, if there is a such a thing, and does not refer, if there is no such thing.

(1_z) "John believes ze inventor of bifocals snored"

and

(2_z) "John believes ze inventor of bifocals snored"

diverge in communicated meaning, and possibly in truth-value, in exactly the same way as (1) and (2).

In general, if we replaced every occurrence of a definite description with a zefinite description, or – better yet – with a zefinite description that is homonymous with the corresponding definite description -- the very facts that motivate the Russell-Frege view would recrudesce. But those facts necessarily give zero support for the view that zefinite descriptions are quantifiers. So those facts give zero support for the view that definite descriptions are quantifiers.

This does not mean, by the way, that Russell's theory has a zero probability of being true. It means only that *those facts* are neutral between it and its negation. Being an empirical theory about natural language, Russell's theory obviously has a greater than zero per cent chance of being true.

These apparent substitution-failures are not so hard to explain; we need only be very clear on what it is that the referentialist really believes. (The "referentialist" is somebody who holds that (tokens of) "the phi" refer(s) to the unique phi.)

Consider:

(1) "John believes the inventor of bifocals snored".

As Russell observed, this has a de re reading and a de dicto reading. The de re reading is:

(2) There is some individual x who uniquely invented bifocals, John believes: that x snored.

The *de dicto* reading is:

(3) John believes: that a unique inventor of bifocals snored.

Russell's theory accommodates the fact that (1) is capable of both readings. But the same is true of the referentialist's theory.

The referentialist says:

If there is an O that uniquely has phi, then "….the phi…" has for its literal meaning the proposition:…O…And if there is no such object, then "…the phi…" has no proposition for its literal meaning; it is abortive.

As we saw earlier, a consequence of the referentialist's position is that "…the phi…" *communicates*, though it does not literally mean, the proposition: *some object O uniquely has phi, and…O…*

So it is actually a consequence of the referentialist's position that:

(1) "the inventor of bifocals snored"

communicates

(5) *some object O uniquely invented bifocals and O snored.*

Given this, if we take (1) and put a "John believes that" in front of it, the result is ambiguous in terms of what is communicated (not what is semantically encoded between). The result is ambiguous between the de dicto and de readings described a moment ago. (1) communicates (5). There are two different places that an operator can be inserted into (5). So putting a "John believes that" in front (1) will be ambiguous (in terms of what is communicated). More exactly, it will be ambiguous between:

(6) *John believes that: for some x, x uniquely invented bifocals, and x snored.*

and

(7) *for some x, x uniquely invented bifocals, and John believes that x snored.*

By exactly similar reasoning, putting a "John believes that" in front of (ii) results in a sentence-token that is ambiguous (in terms of what is communicated) between:

(7) *John believes that: exactly one object O was a first post-master general and O snored.*

and

(8) *Exactly one object O was a first post-master general and John believes that O snored.*

So the referentialist's position actually *explains* the apparent substitution failures that motivate the views of Frege and Russell.

Evans[67] said that, in his view, the theory of descriptions is undoubtedly correct. He came up with an excellent argument for it.

The first person to go into outer space was a Russian named Gagarin. (If I am wrong about this, simply replace "Gagarin" with the name of the actual first cosmonaut.) The sentence

(8) "Gagarin is not Gagarin"

is not only false, but self-contradictory. But the sentence

(9) "Gagarin was not the first person to go into outer space"

is simply false; it is not self-contradictory.

Therefore "Gagarin" doesn't co-refer with "the first man to go into outer-space". Therefore, "the first man to go into outer-space" doesn't refer to anybody.

First of all, there is no denying that what is *communicated* by (ix) is

correct; somebody might have made it into outer space before Gagarin. But let us not hastily make judgments about literal meaning.

As we know, an immediate consequence of the referentialist's position is that: "…the first person to go into outer space…" communicates: *somebody x was uniquely a first person to go into outer-space and…x…*

So an immediate consequence of the referentialist's position is that (ix) communicates, though it doesn't semantically encode, the proposition:

(10) *somebody x was uniquely a first person to go into outer-space and Gagarin was not x.*

(10) is not a self-contradiction, not by anyone's lights. For infinitely many substitutions of the variable, the result is a true proposition. Therefore (10) is a true, and thus non-contradictory, proposition. So the referentialist actually *predicts* that (9) will communicate a non- contradictory proposition. So Evans' argument does not prevail.

Some loose ends

Some hold that definite descriptions are terms that refer to individuals. Within this group, there is disagreement as to what kind of referring terms they are. Some hold that they are indexicals, some do not.

If what we've said is correct, they are indexicals. "The inventor of bifocals" picks out different entities in different worlds. Therefore, that expression is context-sensitive. The semantic content of the type "the inventor of bifocals" is a function from contexts of utterance to individuals. In this case, the contexts happen to be worlds. Given that we never leave our world, definite descriptions seem not to have *as* variable a reference as other indexicals, and thus might *seem* not to be context-sensitive. But they are. Therefore, they are indexicals. This is weakly borne out by the fact that, etymologically, "the" is a truncated form of "that".

Chapter 10 The causal theory of reference

Suppose that, in virtue of having the form "…E…", a sentence-token encodes a proposition that has O as a constituent. In that case, let us say that occurrences of E "semantically contribute" (or, simply, "contribute"). I have repeatedly said that

(SC) E refers to O iff E semantically contributes O.

But (SC) is in apparent contradiction with a very popular theory: the so-called "causal theory" of reference. That theory is:

(CTR) E refers to an object O iff tokens of E have a certain kind of causal relation to O (or states of affairs involving O).[68]

I would suggest that (CTR) is *not* in fact opposed to (SC), at least not in so far as (CTR) is a coherent doctrine.

Some clarifications about CTR are appropriate. First of all, CTR is about expression-*tokens*. The expression-*type* "Socrates" isn't causally connected to anything. Tokens of it have causal relations. CTR says that such tokens refer to Socrates just in case they have a certain kind of causal relation to that man.

There are other, less innocuous, qualifications. No one would maintain that (CTR) is meant to apply to *all* referring terms. Consider "the first animal born in the year 3000 A.D." (If you think definite descriptions are quantifiers, then consider: "ze first animal born in the year 3000 A.D.")

Suppose that Alphie is the animal satisfying that description. In that case, Alphie is the thing referred to by that expression. But "the first animal born in the year 3000 A.D." refers to Alphie by virtue of encoding descriptive information that singles out Alphie; that expression does *not* refer to him by virtue of there being some causal relation between tokens of it and states of affairs involving Alphie. Such a causal connection doesn't exist. Even if it did, it wouldn't be what is operative.

This point can be generalized. (CTR) is not meant to apply to *any* expression that has constituent structure. "That snow is white" refers to the proposition *that snow is white.* The former refers to the latter because the latter is the thing having the properties in the descriptive content encoded in the former. CTR may be meant to apply to the constituents of "that snow is white", but not to that expression as a whole.

So far we've spoken about what CTR does *not* say. To explain what CTR *does* say, we need to define a few terms.

Let R be a referring term belonging to some language L. R is "rigid" iff it refers to the same thing in every world where L is spoken. In every world where English is spoken, "Richard Nixon" refers to Richard Nixon.[69]

Of course, there might be a world where "Richard Nixon" refers to John Kerry, but where *otherwise* English is spoken just as it is here. But, strictly speaking, the language spoken there isn't quite identical with the language spoken here; that version of English comprises the rule:

"Richard Nixon" refers to John Kerry.

Our version of English does not comprise that rule. So we are dealing with different sets of semantic rules, with different languages.

When we talk about rigid designation, we always have in mind worlds where the language in question is characterized by the *exact* same semantic rules that characterize it here.

There are two kinds of rigid designators. "Richard Nixon" refers to the same thing in all worlds where English is spoken. This is a matter of its *semantics.* The semantic rule for that expression is:

(RN) "Richard Nixon" refers to O exactly if O is identical with Richard Nixon.

"Richard Nixon" is what we call a *de jure* rigid designator.[70] Given only its semantics, "Richard Nixon" must refer to Richard Nixon. Nothing non-semantic – not even facts about mathematics or logic – are needed to secure that connection.

"The square of two" is a rigid designator: it refers to the same thing

in every world where English is spoken. But the *semantics* of that expression leave it open what it refers to in a different world. The semantic rule for that expression is:

(s) "the square of two" refers to a number n exactly if n is what results when two is squared.

The semantic rule is *not*:

(s*) "the square of two" refers to four.

In a world where English is spoken, but where the square of two is five, "the square of two" refers to five.

Granted, there probably is no such world. So "the square of two" refers to four in every world where English is spoken. But that is a fact about mathematics: the *semantics* involved leaves it open whether "the square of two" refers to five or not.

"The square of two" is a *de facto* rigid designator. It does denote the same thing in all worlds where English is spoken. But this is a consequence of *non*-semantic facts; by themselves, the semantics of that expression do not guarantee that it will refer to four.

"Four" is a *de jure* rigid designator. The semantic rule for it is:

(FR) "Four" refers to O exactly if O is identical with the number four.

So no matter what laws of arithmetic obtain in a world w, "four" refers to four in that world, so long as English is spoken there; by themselves, the semantics of that expression guarantee that it refers to four.

(CTR) is meant, presumably, to apply to *de jure* rigid designators, like "Nixon" and "Socrates". It is not meant to apply to *de facto* rigid designators like "the number of primes between 10 and 20". *De facto* rigid designators refer by virtue of encoding the right descriptive information, never merely by virtue of a causal connection.

Even with all of these restrictions, problems arise for (CTR). That doctrine has a hard time dealing with expressions that are *de jure* rigid designators, such as "2" and "justice". These terms denote abstract objects.

Nothing has any causal connection to such an object. So it isn't clear how CTR can deal with such terms.

I myself think that the causal theory is wrong even for "Nixon" and "Socrates". But this is neither here nor there. What I wish to show now is that *even if* (CTR) is unrestrictedly true, there is still no opposition between it and (SC). There is no more opposition between those two views than there is between the view that a triangle is a closed, three-sided, planar, straight-edged figure and the view that a triangle is the area bounded by three coplanar lines such that any two of them intersect, but not all three of them intersect.

The right causal relation

The causal theorist must say *which* causal relations constitute, or underwrite, reference. Not just any causal relation will do. If I push the door-bell, that causes the bell to ring. But the ringing of the bell doesn't refer to my pushing of the button.

For the sake of argument, suppose that (CTR) is correct. Let R be the causal relation, whatsoever it should turn out to be, that constitutes or underwrites the relation of reference. In other words, R is the causal connection that causal theorists have in mind (or, better, that they are seeking).

How do we know whether a given type of causal connection is the right one? What is our benchmark? Suppose some not very intelligent causal theorist says:

(*) Consider the causal relation that holds between pushing the button at t and the bell's ringing at t. Let us refer to that relation as DB. R is the kind of causal relation that holds between Socrates (or, more exactly, states of affairs involving him) and tokens of "Socrates" – it is the relation that makes those tokens refer to that person.*

Everyone would *reject* (*). DB obviously does *not* do the job. If indeed tokens of "Socrates" stood in that relation to Socrates himself, that would *not* be enough for such tokens to refer to the man. Why not?

If you want to make a statement *about* some pushing of some button, you cannot use the resulting bell-ringing as a way of doing so. Suppose that the button is pushed at t, and the bell rings at t*. Let RING be that particular ringing, and let PUSH be the button-pushing in question. The sound-token:

(#) "[RING] [71] woke the dog."

does *not* bear the proposition:

(##) PUSH woke the dog.

(#) doesn't say *anything.* You have an unmeaning noise, followed by "is a distal cause of the dog's waking up". So (#) no more bears (##), or any other proposition, than does:

(###) "posidhfoasfhi woke the dog."

What RING is missing is the ability to play a role in sentences.[72] RING cannot be a part of a well-formed English sentence. That fact, by itself, bars us from saying that RING *refers* to anything, at least in English. In order to qualify as an expression of English that refers to PUSH, RING must be able to figure *grammatically* in English sentences. It would be absurd to say that a referring term *belonging to English* could not be used grammatically in English sentences.

An expression can be used grammatically in a sentence of English only if it can make a semantically significant contribution to such a sentence. So if it is to refer to RING, PUSH must be able to make such a contribution.

But that is not enough. The word "Morgan Freeman" makes semantically significant contributions to English sentences. But "Morgan Freeman" doesn't refer to PUSH.

If RING is to refer to PUSH, it is not enough that it make just any semantic contribution; it must contribute some specific thing. The thing it contributes must presumably be PUSH itself.

DB is the wrong causal relation because it does not enable RING to contribute PUSH to the meanings of sentences.

The other side of the coin is that, if DB did enable RING to

contribute PUSH to the meanings of sentences, then DB would be the right causal relation.

No causal theorist holds that any causal relation is reference-constituting. How does he decide which causal relation is the right one? The decision is based on considerations like the ones mentioned. Suppose x stands in causal relation R to y. If that is necessary and sufficient for x's semantically contributing y, then R is the right relation. If not, not. So R is the right causal relation exactly if it is consistent with SC.

A qualification

RING could refer to PUSH. Indeed, RING could refer to PUSH in virtue of standing in a certain causal relation to PUSH. But that would not be enough to make the causal theory be correct.

Let L be a language that is just like English except that, by our stipulation, L is characterized by the following semantic rule. In L, the ringing produced by a door-bell pushing actually *refers* to that doorbell pushing. Let RING* be some particular door-bell ringing. In L, the spoken sentence-token "[RING*][73] woke the dog" means that a certain door-bell pushing woke the dog. [74]

Here we have a clear case where one thing refers to another in virtue of there being a causal connection between the two.

But notice that, by itself, the causal relation does not suffice for the reference relation. What suffices is that causal relation *plus* some semantic rule to the effect that if x and y stand in that causal relation, *then* sentences of the form "…x…" are to mean:…*y*…

The content of the causal theory

This last point gives us some insight into what can legitimately be alleged by a causal theory of reference. This insight will, in turn, substantiate our analysis of reference.

For the sake of argument, suppose that the causal theorist is right; suppose there is some causal relation R such that x (a given sound or inscription) refers to y iff y stands in R to x.

In that case, tokens of "Socrates" refer to Socrates because Socrates

stands in relation R to those tokens. But it is easy to show a causal relation is never sufficient for reference. And when we make it clear what *is* sufficient, it turns out that a causal connection between object and token has a decidedly subordinate role. Obviously there is an Albanian word for Socrates, a Finnish word for Socrates, an Urdu word for Socrates, and so on. These are different words. According to the causal theorist, for any token *t* of any one of these words, what makes *t* refer to Socrates is that Socrates stands in R to *t*. The Urdu word for Socrates is "Sukrat". So according to the causal theorist, tokens of "Sukrat" refer to Socrates because Socrates stands in R to "Sukrat". The same is true *mutatis mutandis* of the Albanian word for Socrates, the Finnish word, the Hungarian word, and so on.

Let us focus on "Sukrat." If you were to put "Sukrat" in an English sentence (I mean, if you were to *use* N in an English sentence, not *mention* it), the result would be ungrammatical; you would not have a well-formed sentence of English. If you were to say "Sukrat was wise" or "Sukrat drank hemlock", the result would not be a well-formed sentence. People might understand you – after all, communication doesn't always involve well-formed utterances. (If I child emits a certain kind of cry, I know that it wants food or wants its diaper to be changed. But its cry is not English or Urdu or any other language.) When you said "Sukrat was wise", they might say: "You mean *Socrates* was wise! I didn't understand you at first." Their understanding you would involve their doing a bit of translation. This means that your utterance was not a well-formed utterance of *English*; it contained a renegade constituent, one that had to be replaced with an English word to yield proper English.

This point is liable to be misunderstood; maybe a different example would help. The Spanish word for "wet" is "mojado". According to the causal theorist, tokens of "mojado" refers to the property of wetness exactly because they stand in a certain causal relation to instances of wetness. If you were to say "Smith is mojado", what you were saying would be patently ungrammatical. The same thing is true of "Sukrat", though perhaps less obviously. If you said "Sukrat was wise", what you were saying would be ungrammatical. If you resist this, then (mutatis mutandis) simply replace "Sukrat" with "mojado" in the present discussion.

Of course, what is ungrammatical today might be grammatical tomorrow. Languages are plastic entities, at least along certain dimensions. Lexical items can be added quite freely. (Non-lexical changes – changes in

syntax or grammar – are less easily made and are virtually never made intentionally.) So if you had a friend who always said "Sukrat" instead of Socrates, you would, in effect, *add* a rule to your idiolect to the effect that "Sukrat" refers to Socrates. But until such a rule were added to English – and, as of yet, it hasn't been added – "Sukrat" is not in the English lexicon; it is not given a meaning by the semantics of English; therefore "Sukrat was wise", though perhaps easily enough understood, is not well-formed; it is not grammatical.

Given any referring term, it will be a referring term *of* this or that language. It will not be a referring term *simpliciter*. "Sukrat" refers to Socrates in Urdu, not in English. "Socrates" refers to Socrates in English, not in Urdu.

If the causal theory is right, then what makes tokens of "Sukrat" refer to Socrates is wholly that Socrates stands in relation R to those tokens. After all, the causal theorist says that x refers to y iff y stands in relation R to x. Tokens of "Sukrat" stand in relation R to Socrates. Therefore such tokens refer to Socrates.

But *in English* "Sukrat" obviously does *not* refer to Socrates. Quite possibly, in any language other than Urdu, "Sukrat" does not refer to Socrates.

But the causal theorist says that what makes tokens of "Socrates" refer to Socrates is simply that the latter stands in R to the former; the reference-relation in question is *constituted by* that causal relation. That causal relation is quite enough, according to the causal theorist, for "Socrates" to refer to Socrates in English. But the problem is that tokens of "Sukrat" stand in that relation to Socrates no less than tokens of "Socrates". And yet tokens of "Sukrat" clearly *do not* refer to Socrates in English. So for x to refer to y in some language L, it is *not* enough that y stand in R to x. That may be a part of the story; but it cannot be the whole of it.

The causal theorist might resist this, and say:

The causal relation that an Urdu word must bear to Socrates, if that word is to refer to him, is different from the causal relation that an English word must bear to Socrates, if that word is to refer to him. Both of those relations are different from the causal relation that a Tagalog word must be to Socrates, if that word is to refer to Socrates. Different causal relations are involved in different languages.

This position disjunctivizes the concept of reference. There is referenceinUrdu, referenceinEnglish, referenceinFrench, and so on.[75] But there is no longer a *single* concept of reference. This is plainly unacceptable. "Sukrat" refers to Socrates in the very sense in which "Socrates" refers to him, and in which the Chinese word for Socrates refers to him.

To avoid disjunctivizing the concept of reference, the causal theorist must say that some *one* relation R does the job. But we have already seen why such a position is untenable.

At this point, the causal theorist may weaken his position and say:

I never said that reference *is* some causal relation; I said that a certain causal relation is essentially *involved* in reference. So reference *per se* is not some causal relation; it is something else. That something else *involves* some kind of causal relation.

But then we are left with *no* answer to the question: "what is reference?" Yet this is the very question that CTR is supposed to answer.

A place for CTR

Still, there is a certain truth in CTR. Consider the English semantic rule that is involved in the fact that "Socrates" refers to Socrates. That rule may well *exploit* some causal connection obtaining between certain current sound- and inscription-tokenings, on the one hand, and states of affairs involving Socrates, on the other. That rule may be at least vaguely like:

(CRS) Let s be an arbitrary token of the physical type "Socrates". If there is some object x such that x uniquely stands in causal relation R [for the appropriate value of R] with respect to s, then s refers to x. If there is no such object, then s refers to nothing. If there are many such objects, then s is ambiguous between them.

So given (CRS), and given that Socrates is the thing to which my token of "Socrates" stands in relation R, it follows that Socrates is what that token refers to. If Plato were the thing to which that tokening stood in that relation, then that tokening would refer to Plato.

In general, causal connections are involved in reference only when semantic rules exploit them.

This insight enables us to circumvent some problems associated with CTR. We can obviously refer to things to which we have no causal connection – for example, future individuals and abstract objects.

But there is only one concept of reference. "2" refers to the number two in exactly the sense in which "Nixon" refers to Nixon. There is no causal connection between the first pair. So CTR disjunctivizes the concept of reference.

Here we are dealing with a *different* disjunctivization of reference from the one discussed before. So CTR is guilty of doubly disjunctivizing an obviously unified concept. Thus CTR, if taken as a thesis as to what *constitutes* reference, is doubly wrong.

Advocates of CTR have tried to work around this problem, saying things like: *Tokens of "2" are causally connected to instances of the number two; and that is why such tokens refer to that number.*[76]

But such maneuvers are obviously implausible. One can refer to the number two *itself*, not just to instances of the entity; and we have no causal connections to that entity itself.

Further, causal connections aren't fine-grained enough to constitute-reference: I can refer to the clay, the statue, time-slices of the statue. I am referring to a different thing in each case. But any causal connection I have to the one is equally a causal connection to the others.[77]

Even if these problems can be finessed, CTR cannot possibly be true of referring terms that have phrasal structure, like "that Socrates wept often" and "the [ze] center of mass of the galaxy". Such terms refer by virtue of describing, not by virtue of a casual connection. If CTR says that causal connections *constitute* reference, instead of being merely involved, then it says that the sense in which "Nixon" refers to Nixon is entirely different from the sense in which "the phi" (or "ze phi") refers to the phi.

These problems vanish when CTR is put in its proper context.

In some cases, we define a word by *pointing* to its referent: "*that* man over there is Thaddeus Smith." Reference is secured through ostension.

This is not always how words are defined. When we leave aside ostensive definition, we find that reference is always secured in the same way. We *describe* the thing to which an expression is meant to apply. But we give the description wide-scope. That way the sole semantic content of the

expression is the individual: the expression doesn't become a Russellian description-quantifier.

Suppose I want to produce a term that refers to the first individual born in the 3rd millennium. I say:

(3M) *If* there is some object O such that O is the first individual born in the 3rd Millennium, then "Simon" refers to O.

Because the descriptive content in that rule is given wide-scope with respect to the "refers to" operator, the statement "There are possible worlds where Simon is born in the 2nd millennium" is true. This is what we want: after all, any individual *could* have had a slightly different birth-date.

As we've seen, in so far as CTR has any chance of being right, it is because certain semantic rules *exploit* causal connections. If a causal connection between tokens of "Socrates" and Socrates is involved in the fact that the former refer to the latter, it is because some rule like the following exists:

(S) If there is some individual O such that tokens of "Socrates" stand in R to O, then those tokens refer to O.

(S) and (3M) are completely parallel. In both cases, reference is secured through descriptive information. But that information is given wide-scope with respect to the "refers to" operator. Where (S) is concerned, the descriptive content has to do with causal relations; where (3M) is concerned, this is not so. But that doesn't have any special ramifications. Descriptions differ; some involve the concept of causality; some don't. Some involve the concept of wearing a feathered-hat; some don't. I can describe an entity in terms of its causal liaisons or in terms of what kinds of clothes it is wearing. CTR merely registers the banal fact that, in some cases, the descriptive information in question has to do with a thing's causal liaisons.

Once this is realized, we don't have to invent tortured hypotheses as to why there really is a causal connection between the number two and tokens of the expression "2". There is no such connection; such tokens are not

causally connected to the platonic type *two*, but only to its instances. Even if there were such a causal connection, it wouldn't be reference-constituting. What would be reference-constituting would be a semantic rule that fixed the referent of "2" in terms of descriptive information. That rule might exploit a description that alluded to the causal liaisons of certain things (although, in actuality, it would not). But that doesn't warrant a causal theory of reference. If I say "Jim is the man wearing the feathered hat", I am exploiting the fact that Jim is wearing a certain hat. In effect, I am saying: "there is some contextually salient object O that is wearing a feathered hat: 'Jim' refers to O." But this doesn't warrant a feathered-hat theory of reference.[78] [79]

A nicety

Uncontroversially, it is *possible* that "Sukrat" refers to Socrates only in Urdu. We've seen how this fact causes trouble for the causal theory.

In my view, it is *analytic* that "Sukrat" can refer to Socrates only in Urdu. Given any word w of any language L, it is analytic that w cannot be used grammatically in a sentence not belonging to L. A homonym of w can be so used, but not w itself. Perhaps a homonym of "Sukrat" refers to Socrates in some language other than Urdu. In that case, that homonym can be used grammatically in some language other than Urdu. But the *Urdu* word "Sukrat" cannot refer to Socrates, or have any grammatically licensed role, in any language other than Urdu.

Suppose the gases brewing in a volcano yield the noise "Socrates was wise". The volcano hasn't really said anything. It produced sounds that *would* have meant something if *we* had produced them. When we say "Socrates was wise", it is in consequence of our following certain semantic rules. If we were following no such rules, then we wouldn't be using a language at all; we wouldn't be saying anything. By exactly similar reasoning, what makes a sound be a token of a word is that one produces that token in consequence of one's following a certain semantic rule; if one is just *caused* to blurt it out – the way a corpse might be electrically stimulated to make noise – one wouldn't really have produced a bit of language at all. Unless we have something that is a consequence of one's awareness of a semantic rule, we don't have a word-token.

A corollary is that *which* word you are tokening is not determined solely by the phonetic properties of your utterance. If you really produced a

word-token, and not just a noise that sounds like a word-token, that means you were following a semantic rule, and that you produced that word-token as a consequence thereof. Which word you tokened is a function of which rule you were following.

As we said before, a word never refers *simpliciter*; it always refers in this or that particular language. The semantic rule for "Socrates" is not: "Socrates" refers to Socrates. It is: "Socrates" refers to Socrates *in English.*

Suppose a speaker of L produced the sound "Sukrat" in consequence of following the semantic rule of *L* that assigns Socrates to "Sukrat". What he has tokened is a word of L, not of Urdu. He was not following an Urdu semantic rule, and thus didn't token an Urdu word. So strictly speaking, his utterance was ungrammatical, even though that ungrammaticality is completely undetectable.

§ Here is Frege's position. Referring terms have both "sense" and "reference". The sense is a concept; the referent is the unique entity falling under that concept. Consider "the inventor of bifocals". The sense is a concept -- *unique inventor of bifocals.* The referent is the entity satisfying that concept, namely Franklin. Now consider "the first post-master general". The referent is the same as before, but the sense is different. In this case the sense is the concept *unique first post-master general.*

It is because of these differences in sense that the following two sentences convey very different propositions:

(i) "The inventor of bifocals snored."
(ii) "The first post-master general snored".

In both cases, the same property is being attributed to the same person. But different senses or concepts are being used to pick out that person. As a result, the information that (i) gives you is different from the information that (ii) gives you. In the one case you are learning that a unique bifocal inventor snored; in the other case, you are learning that a unique first post-master general snored. So what is learned in the one case is very different from what is learned in the other.

There are obviously elements of deep truth in Frege's position. But

there are also deep muddles. To see the truth in Frege's position, we must first identify the muddles.

First, there is no such thing as *the* expression "the inventor of bifocals" or *the* sentence "the inventor of bifocals snored". There are expression-types and expression-tokens. Type-meaning is different from token meaning. The meaning of an expression-type is a rule saying how to compute the meanings of the corresponding tokens. The meaning of the *type* corresponding to (i) is a rule saying which propositions are meant by tokens of that type. The meaning of one of those tokens is the proposition assigned it by that rule. The meaning of the *type* "the inventor of bifocals" is a rule saying what its tokens refer to. The meaning of a given token is the thing referred to.

The rule that assigns propositions to tokens of (i) is very different from the rule that assigns propositions to tokens of (ii).

In the first case, the rule is this. *if* somebody x uniquely invented bifocals, then tokens of "the inventor of bifocals snored" are true exactly if x snored. Let R1 be this rule.

In the other case, the rule is this: *if* somebody x was a unique first post-master general, then tokens of "the first post-master general snored" are true exactly if x snored. Let R2 be this rule.

To understand an expression is to be able to assign it the right meaning. Doing this involves applying the relevant semantic rules.

To understand a token of (i), one must apply R1. Inevitably, to anyone who understands such a token, it will communicate (inter alia) that somebody x uniquely invented bifocals and x snored.

To understand a token of (ii), one must apply R2. Inevitably, to anyone who understands such a token, it will communicate (inter alia) that somebody x was a unique first post-master general, and x snored.

So what a token of (i) communicates is different from what a token of (ii) communicates. But this is not because they encode different propositions. It is because the relevant semantic rules are different.

R1 gives the semantics of the *type* corresponding to (i), and R2 gives the semantics of the type corresponding to (ii) .So the semantics of the *type* corresponding to (i) is different from the semantics of the *type* corresponding to (ii). This is equivalent to the point we made a moment ago: the rule that assigns meaning to tokens of the one type is different from the rule that assigns meaning to tokens of the other type.

Thus, the differences in cognitive value between tokens of (i) and tokens of (ii) have to do with differences in the semantics of the corresponding *types.* Those differences do not have to do with semantics of the tokens themselves.

Where definite descriptions are concerned, what Frege calls "sense" is type-semantics, and what he calls "reference" is token-semantics. The semantics of the type "the inventor of bifocals" is given by the rule: for any x, a token of "the inventor of bifocals" refers to x exactly if x falls under the concept *unique inventor of bifocals*. The semantics of such a token is the thing assigned it by the rule just given – it is the referent of that token.

It is expression-tokens that refer. Expression-types *assign* referents to their tokens, but they do not themselves refer. In virtue of having the form *the inventor of bifocals has phi*, a sentence-token encodes the proposition *Franklin has phi.* The occurrence of "the inventor of bifocals" is obviously associated with the concept *inventor of bifocals*. But that association lies in the semantics of the corresponding type. The semantic rule for that *type* is: any token of "the inventor of bifocals" refers to x exactly if x uniquely satisfies the concept *inventor of bifocals.*

Frege thought that, in virtue of having the form *the inventor of bifocals has phi*, a sentence-token encodes a proposition that is true exactly if there is a unique bifocal inventor x, and x has phi. This view is a mistake. The concept *inventor of bifocals* does not make it into the proposition meant by tokens of *the inventor of bifocals has phi*. That concept is part of the rule that *picks out* a thing which is to be a constituent of that proposition. But that concept is not itself such a constituent. As we saw earlier, if that concept *were* such a constituent, then "the inventor of bifocals" would be a quantifier, not a singular term.

But, as we've seen, even though that concept is not itself such a constituent, its presence in the relevant type-semantic rule is enough to ensure that such a token will *communicate* the proposition: *exactly one thing x invented bifocals, and x has phi.*

The distinction between sense and reference is obviously legitimate, and it obviously does the job that Frege wanted it to do: it accounts for the differences in cognitive value between utterances of (i) and utterances of (ii). What Frege didn't realize is that sense is type semantics and that reference is token semantics. This failure led Frege to have an erroneous view regarding literal meaning, and a related erroneous view regarding the nature of

reference. The erroneous view regarding literal meaning is this: what is literally meant by occurrences of (i) is some proposition that has for a constituent, not Franklin himself, but some concept that singles out Franklin (namely, *inventor of bifocals*). The erroneous view regarding reference is this: if e refers to o, then in virtue of having the form "…e…", a sentence encodes a proposition that has, not o itself as a constituent, but some *sense* that singles out o.

I believe that Kaplan had considerations like this in mind when he wrote[80]:

"Fregean sinn conflates elements of two quite different notions of meaning. One, which I called character, is closed to the idea of linguistic meaning (and perhaps of cognitive content). Another, which I call content, is what is expressed by an expression in a particular context of use. The content of an utterance of a complete sentence [sentence-token] is a truth-bearing proposition. Where indexicals are involved, the difference between character and content is quite clear."

The point Kaplan is making here is profound and correct. But an addition must be made to it. We saw earlier that there are two kinds of pre-semantic implicature. There is the kind associated with type-semantics – with the rules that *assign* meaning to expression-tokens. But there is also the kind that is associated not with type-semantics itself, but with one's *access* to type-semantics. Frege held that *proper names* have both sense and reference: he didn't confine himself to definite descriptions. Where *definite descriptions* are concerned, sense is indeed type-semantics, as Kaplan says. But where *proper names* are concern, sense is *not* type-semantics: it is rather the information through which one *grasps* type-semantics.

Frege thought that, at the level of literal meaning, proper names ("Hesperus" and "Socrates") are really definite descriptions ("the last celestial body to disappear from the morning sky", "the great philosopher who died of hemlock-poisoning"). Saul Kripke (1972, 1979) refuted this: proper names have no sense – they are just labels. (Attempts have to re-descriptivize names. But Soames (2001, 2005) did a good job of shutting down these attempts.)

But Kripke's important semantic points leave a major question unanswered. In fact, they don't deal with any of the data that led Frege to regard names as descriptions. Suppose "Hesperus" and "Phosphorous" are mere labels. In that case, they label the same thing, and those two sentences have exactly the same literal meaning. There can be no doubt that

(a) "Hesperus is lovely"

and

(b) "Phosphorous is lovely"

convey different propositions.

How are we to deal with this? Frege answered by saying that "Hesperus" and "Phosphorous" are *not* mere labels. Kripke refuted that answer. But Kripke does not himself say how we are to deal with the problem.

There is more to say. The proposition *conveyed* by occurrences of (i) has as a constituent the concept *last celestial body to disappear from the morning sky.* And the proposition *conveyed* by occurrences of (ii) has as a constituent the concept *first celestial body to appear in the evening sky*. Those propositions are exactly what Frege thought.

So even though "Hesperus" and "Phosphorous" do not have senses *at the level of semantics*, they do have senses *at the level of implicature.* What exactly is responsible for that difference in implicature? Here we *cannot* say what we said in connection with definite descriptions. Tokens of

(i) "The inventor of bifocals snored."

differ in implicature from tokens of

(ii) "The first post-master general snored".

Here the difference lies in facts about type-semantics. The rule assigning meaning to tokens of (i) involves the concept *unique inventor of*

bifocals, and does not involve the concept *first post-master general*. It is the other way around as regards the rule assigning meaning to tokens of (ii).

But there is no comparable difference where the type-semantics of (a) and (b) are concerned. There is some object O such that the rule assigning meaning to tokens of (a) is:

(ar) Tokens "Hesperus is lovely" are true exactly if O is lovely.

And there is some object O* such that the rule assigning meaning to tokens of (b) has the form:

(br) Tokens "Phosphorous is lovely" are true exactly if O* is lovely.

Moreover, O is identical with O*. So (ar) and (br) are nearly enough same rule. The only difference is that where one of them has "Hesperus" the other has "Phosphorous". *Neither* rule includes the concept *last celestial body to disappear from the morning sky* or the concept *first celestial body to appear in the evening sky*. The just mentioned difference between those two rules is far too trivial to ground the massive differences in implicature between tokens of (a) and tokens of (b). The cognitive difference between a token of (a) and a token of (b) is not merely verbal; it isn't *merely* that, where one of them uses one expression, the other uses a different expression. The difference isn't comparable to that between "Fred is an enemy" and "Fred is a foe". So given only the type-semantic differences corresponding to (a) and (b), there is no prospect of explaining the cognitive differences between tokens of the one and tokens of the other: the strategy we applied to definite descriptions won't work here.

In this case, the differences in cognitive value are to be explained not in terms of type-semantics, but in terms of our access to type-semantics.

Let us start by discussing the semantic rules associated with "Hesperus" and "Phosphorous". There is some object O such that the semantic rule for the type "Hesperus" is this:

(hr) Tokens of "Hesperus" refer to O.

And there is some object O* such that the semantic rule for the type "Phosphorous" is this:

(pr) Tokens of "Phosphorous" refer to O*.

Since O=O*, (hr)=(pr).

Let us focus on (hr). To *grasp* (hr), you must have some way of singling out Hesperus. Otherwise, you will, at most, grasp something like (hr) that has a gap in it where that object is supposed to be. You do not grasp objects in a vacuum. For reasons discussed earlier, knowledge of descriptive information is involved in the grasping of any external object. So your *grasp* of (hr) will have descriptive content that (hr) itself does not.

Let us make these points more concrete by considering one way that you learn the semantic rule for "Hesperus". Somebody says to you:

(A) "'Hesperus' is the name of the last celestial body to disappear from the morning sky".

Here what you are being told is:

(B) Something x is a unique last celestial body to disappear from the morning sky, and the semantic rule for "Hesperus" is: "Hesperus" refers to x.

What you are being told is *not*:

(C) The semantic rule for "Hesperus" is: something x is a unique last celestial body to disappear from the morning sky "Hesperus" refers to x.

And what you are being told is *not*:

(D) The semantic rule for "Hesperus" is: "Hesperus" refers to anything x that is a unique last celestial body to disappear from the morning sky

The semantic rule for "Hesperus" doesn't have anything to do with the concept *last celestial body to disappear from the morning sky.* There are possible worlds where the type "Hesperus", as well as its tokens, have the same semantics as their counterparts here, but where Hesperus is *not* a unique celestial body to disappear from the morning sky. (B) is consistent with this fact. (C) and (D) are not consistent with it.

(B) gives appropriately narrow scope to the concept *semantic rule*. Because it has appropriately narrow scope, the semantic rule in question is kept pure: the concept *last celestial body to disappear from the morning sky* is no part of it.

(C) and (D) give the concept *semantic rule* inappropriately wide-scope. As a result, the concepts *evening sky*, and so forth, contaminate the semantic rule for "Hesperus"; so (C) and (D) give the wrong semantic rule.

In every world where English is spoken, tokens of "Hesperus" refer to Hesperus. (B) is consistent with this fact. (C) and (D) are not.

People often learn the semantic rules for proper names through statements like (A). Since they learn the *right* rules, we must assume that they give appropriately wide scope to the descriptive information and appropriately narrow scope to the semantic operators.

You do not learn the semantic rule for "Hesperus" in a vacuum. It must be given to you through some statement like (A): that rule must be *described* to you. The rule itself is descriptively innocent; the concept *last celestial body to disappear from the morning sky* doesn't make it in there. But the *description* of that rule does include that concept. So your *access* to that rule is mediated through that concept.

What tokens of *Hesperus has phi* convey to you will reflect the descriptive information through which you access their literal meanings. Suppose you are given the semantics of the type "Hesperus" through some definition like (A). In that case, you access that semantics through (B). Given this, what a token of *Hesperus has phi* will convey to you will be much more than some bare, object-dependent proposition: it will convey to you (inter alia) that something x is uniquely a last celestial body to disappear from the morning sky and that x has phi. At the same time, given that (B) is correct, there is some object O such that the literal meaning of such a token is simply: *O has phi.*

Now suppose you are given the semantics of "Phosphorous" through the following statement:

(D) "Phosphorous" is the name of the first celestial body to appear in the evening sky.

For reasons exactly similar to those just given, what a token of *Phosphorus has phi* will convey to you will be (inter alia) that something x is uniquely a first celestial body to appear from the evening sky and that x has phi. At the same time, there is some object O such that the literal meaning of such a token is simply: *O has phi*.

There is more to say. Any competent speaker would take (A) in the right way; that person would give the amount of scope to the relevant operators. As a result, that person will have the right *modal intuitions.* That person will not think it *necessary* that anything be a unique last celestial body to disappear from the morning sky. And that person will not regard as analytic – as holding merely in virtue of linguistic rules -- the statement "Hesperus is the last celestial body to disappear from the morning sky". After all, that person gave appropriately narrow scope to the concept *semantic rule* – that person took (A) to mean (B), and not (C) or (D). So that person knows that the *semantics* of "Hesperus" do not require that tokens of "Hesperus" refer to a last celestial body to disappear from the morning sky. Exactly similar remarks apply to (D). So that person will have all the right intuitions about modality and analyticity. But at the same time, for the reasons discussed a moment ago, what tokens *Hesperus has phi* convey to her will be very different from what is conveyed by tokens of *Phosphorus has phi*. Moreover, what is conveyed by such tokens will correspond to what Frege saw as the *literal* meanings of those utterances. As we just saw, what is conveyed by those tokens doesn't correspond to type-meaning: it corresponds to one's *access* to type-meaning. The concept last celestial body to disappear from the morning sky isn't part of the semantics of the type "Hesperus"; it is, at most, a part of one's *access* to the semantics of that type. Exactly similar remarks apply to *every* proper name. Where any proper name is concerned, Fregean sense corresponds not to type-semantics, but to one's access to type-semantics. So, where proper names are concerned, Frege's conception of "sense" is not a conflation of token-meaning with type-meaning, but of token-meaning with the information through which one grasps type-meaning. If we think of type-meaning as "pre-semantics" – the semantic content which assigns semantic contents to tokens – then the information through which one

grasps type-meaning is *pre*-pre-semantics: it is the information which enables one to grasp the information which, in its turn, enables one to compute token-meaning. So Frege's view that proper names have senses is a conflation not of semantics (token meaning) with pre-semantics (type-meaning), but of semantics with pre-pre-semantics.

§ Here we must discuss an ambiguity in the term "meaning". At first, this ambiguity might seem to bear against our view that, where devices of reference are concerned, there is no knowledge of meaning without knowledge of referent.

Let us tell a story. Your friend Smith is a Finn who does not speak English. But you speak both Finnish and English. One day, you and Smith hear :

(*) "If penguins liked classical music, and they were analytical as opposed to emotional in temperament, then they would prefer Beethoven to Chopin."

(Suppose that (*) is a particular utterance.) Of course, *you* understand this utterance perfectly well. But Smith does not. But you *tell* Smith that (*) encodes the proposition:

(**) *If penguins liked classical music, and they were analytical as opposed to emotional in temperament, then they would prefer Beethoven to Chopin.*

So now Smith knows exactly which proposition is encoded in (*). Question: Does Smith *understand* what is meant by (*)? The obvious answer is: "yes – to understand an utterance is to know which proposition it encodes; and this is exactly what Smith knows in the case in question." But there is plainly a sense in which you do, and Smith does *not*, (*). You know the "derivation tree" for that utterance. You know not only that (*) encodes (**); you also know *how* (*) it encodes (**). You know that "penguins" means *penguins*; you even know that the "s" in that word indicates plurality; you know that "would" indicates the conditional mood; that "prefer" refers to a certain propositional attitude; and so on. Smith doesn't know of any of this. Really, what Smith knows is *not* that (*) means (**); what he knows is that:

(***)
"ifpenguinslikedclassicalmusicandtheywereanalyticalasopposedtoemotionalintemperamentthentheywouldprefertobeethoventobhopin."

means (**).

Where you hear articulations, Smith hears mere noise. Smith knows that (***) *as a whole* encodes (**), but he doesn't know how this encoding is done exactly. Speaking not very accurately, we might say that Smith doesn't know the right "derivation tree".

Here is the moral of the story. To know the meaning of a sentence-level utterance is *not* to know which proposition is encoded in it. It is to know *how* hook up that utterance to the right proposition. To understand is to know how to use linguistic rules to *find* the right proposition. If you don't have this knowledge, then all you have is the pseudo-understanding that Smith has.

Here is a parallel. (I am going to speak very elliptically, since the nuances are not relevant.) There is a big difference between knowing *that* Fermat's theorem is true, and knowing *why* it is true. I can learn the former by being *told* as much by a reliable mathematician: personal testimony is enough – no mathematical acumen is needed. But I cannot possibly understand *why* it is true unless I am able to follow the relevant proof. So understanding consists in knowledge of derivations, not in knowledge of theorems. So far as (*) is concerned, what Smith is knowledge of a theorem, not knowledge of a derivation. (This is *literally* the case – at least if we conceive of languages as calculi. If that is the conception, then all derived semantic rules are theorems; and any act of computing literal meaning is the act of deriving or proving a theorem. For the record, this *is* how I think of languages.)

Let us encode this in the following slogan (it is understood that the "meanings" referred to are those associated with sentence-level utterances):

(KDT) "Knowledge of meaning= knowledge of derivation tree."

(KDT) is to be distinguished from

(KP) "Knowledge of meaning=knowledge of proposition alone."

Supposedly, we have just seen that (KP) is false and that (KDT) is true.

Here is where the problem for our position begins. The position we *rejected* but that is intuitively very reasonable is this:

Some evil person stole Jones' wallet. Jones says to me: "the man who stole my wallet is not a person of high morality". I don't know who the thief in question is. So I don't know to whom "the man who stole my wallet" refers. But I obviously understand what Jones said. And it is also pretty clear that "the man who stole my wallet" refers to somebody – it refers to the man who stole my wallet.

Here is what we said (it is similar to what Russell said many times). *If* the utterance of "the man who stole my wallet" really is a referring term, as opposed to a quantifier, then for some individual O, Jones' utterance has for its literal meaning the proposition: *O stole my [Jones'] wallet.* So *if* the definite description refers to somebody, then you *don't* understand that token unless you know that it was (say) *Brown* who stole the wallet.

It might be said that our position presupposes (KP) – the idea that (for sentence-level utterances) to know an expression means is to know what proposition it encodes. But the right principle, supposedly, is (KDT) – to know the meaning of a sentence-level (or otherwise complex) expression is to know its derivation tree. So to understand the utterance in question of "the man who stole my wallet is not a person of high morality" is to know *how* to assign it the right proposition, not *which* proposition it encodes. This applies, in particular, to the definite description. To understand it is not to know *which* thing it refers to; it is to know *how* to "hook it up" to the right thing by manipulating the relevant semantic rules.

So a more refined conception of linguistic understanding bears out the intuition – whose correctness we denied – that one can understand a token of *the phi has psi* without knowing who the phi is.

Here is what I say to this. I agree that true understanding involves knowledge of the right "derivation tree". But knowledge of the right derivation-tree relates to understanding the relevant *type*. (KDT) is true of *type*-understanding. (KP) is true of *token*-understanding.

There is the proposition meant by (*). And then there is a

proposition saying how to compute the proposition meant by (*). The latter is the proposition that gives the "derivation tree" for (*). Consider the sentence-type corresponding to (*). To know the meaning of that sentence-type is to know how to compute, or derive, the meanings of its tokens. So knowledge of derivation-trees is type-knowledge. To know the meaning of the token is to know *which* proposition lies at the end of the derivation tree.

Let us sum up. There is indeed a sense in which Smith does not understand (*). But what he doesn't understand is the corresponding type – he doesn't know the rules which *assign* meaning to (*). But, thanks to your testimony, he *does* know the meaning of (*). So nothing that we've seen in connection with (*) warrants any rejection of the view that to understand a token is to know which proposition it means.

Thus our defense of (NKT) remains intact. As we saw, if a token of *the phi* refers to O, then a token t of *the phi has psi* encodes the proposition *O has psi.* If the concept *phi* were to make it into that proposition, then *the phi* would be a quantifier. So if you don't know that t encodes the proposition *O has phi*, then you don't know what it means. And if you don't know that the occurrence, in t, of *the phi* refers to O, then you don't know what that occurrence means. Once again – where referring terms are concerned, there is no knowledge of meaning without knowledge of referent.

§ A failure to distinguish between what we just called "pre-semantics" and "pre-pre-semantics" lies at the root of the failure on the part of semanticists to account for the cognitive values had by sentence-tokens containing indexicals.

Given what Kaplan (1989) argues, it seems pretty clear that tokens of demonstratives are "directly referential". If Fred says "I am tall", his utterance literally means: *Fred is tall.* If somebody says "that man is tall", and Fred is the man in question, that utterance literally means: *Fred is tall.* If Fred says "I am that man", and his statement is true, then his utterance literally means: *Fred is Fred.* Suppose Fred then says: "I am not that man." In that case, the literal meaning of his utterance is a contradiction: *Fred is not Fred.*

Kaplan's arguments for his position are compelling – I myself accept his position as well his arguments for them. But as we just saw, if Kaplan is right, then what indexical sentence-tokens communicate is extremely remote from what they literally mean. As we saw earlier in connection with both

names and definite descriptions, this by itself doesn't even constitute prima facie evidence that Kaplan's position is wrong. But it is still worth knowing *why* the propositions communicated by indexical sentence-tokens are so different from the ones they literally mean; and it is also worth knowing what it is exactly that such tokens communicate.

Let us start with some background. Kaplan distinguishes between an indexical's "character" and its "content". Character is a feature of indexical *types* and content is a feature of indexical tokens. The character of an indexical is a rule that assigns referents to its tokens. For example, the character of "I" is some rule like: if x is speaking, then a token of "I" refers to x. What character calls the "content" of an indexical token is the thing which the corresponding type assigns to that particular token. So if Fred says "I", the content of that token is Fred himself. What Kaplan calls "character" is clearly a kind of type-semantics', and what he calls "content" is a kind of token-semantics.

Kaplan extends the notions of character and content to whole sentences. The character of the type "I am tired" is a rule that assigns propositions to each token of that type. (In this case, the character would be a rule like: If, at time t, so and so says "I am tired", then that token has for its literal meaning the proposition: *so and so is tired at t.*)The content of a particular token is the proposition that the corresponding character associates with it. So if Fred says "I am tired" at t, then the content of that token is*: Fred is tired at t.*

An utterance of "I am tall" and "that man is tall" have different cognitive values even if, in both cases, the indexical refers to Fred. The obvious way to deal with this is to say that the cognitive differences result from the differences in the *characters* associated with the indexicals.

How exactly is this supposed to work? Here we can tell a story similar to ones told earlier. When you hear an utterance of "I am tall", you don't automatically know its meaning; you have to compute its meaning on the basis of the relevant semantic rules – on the basis of type-semantics. In this case, the rule is: if so and so says "I am tired" at time t, then that token encodes the proposition *so and so is tired at t.* So anyone who knows that rule will know that a given utterance of "I am tired" is true exactly if there is a unique producer x of those sounds, and x is tired. So such an utterance *conveys* an existential proposition, even though it doesn't semantically encode one. The same story applies to "that man is tired", except that, in this

case, the rule is: if in context C, so and so is a salient man, then a token of "that man is tired" at t encodes the proposition: *so and so is tired at t.* So such an utterance *conveys* an existential proposition, even though it doesn't semantically encode one. The existential proposition conveyed by the one token is very different from that conveyed by the other: neither entails the other. So, it seems, when we take character (type-semantics) into account, we have an explanation of the deviations between cognitive and literal meaning.

Here is my view. The character-content distinction is obviously *part* of the story: what was just said is correct *so far as it goes*. But it is incomplete. But there are a couple of reasons why that distinction cannot, by itself, provide a general explanation of the cognitive significances of indexical-tokens and, in fact, why it cannot even provide a complete explanation of the cognitive significance of a *single* indexical-token.

Suppose at t there is some man wearing a disguise who is acting very strangely. You say "that individual is insane". As it happens, the individual in question is Smith. Later, you bump into Smith at a party. He is acting very sanely. You say "that individual is sane." According to Kaplan, the first utterance encodes the proposition: *Smith is insane*, and the second encodes the proposition *Smith is sane.* So if we focus on literal meaning, you've contradicted yourself. But obviously you haven't *really* contradicted yourself: at the level of what is *implicated* (of what is communicated but not literally meant), what you've said is consistent. You are not in the same category as somebody who has said "that man is both sane and insane". Here we cannot use the content-character distinction to any advantage. In both cases, the content of the indexical is the same (*Smith*), and so is the character – after all, the same indexical was tokened on both occasions.

To deal with this case, we have to go beyond semantics – we must do a little epistemology. As we said earlier, any sense-perception is descriptive. You don't *just* see Smith; you see *a* thing with such and such properties. Roughly, you see states of affairs, not things. Given this, let us consider the story just told, focusing on your utterance of "that man is insane". You were seeing Smith. You weren't *just* seeing Smith. You were seeing *a* man dressed like a pirate, jabbering incoherently, brandishing a saber, and so forth. And you didn't just see *a* man with those properties; you saw him as being in a certain place – he is *over there,* next to that beige Buick. Now consider the person to whom you directed your utterance of "that man is insane". What we just said about what you saw applies to that person

(give or take some niceties relating to the fact that he wasn't standing in *exactly* the same place as you). So the proposition communicated to your companion was not *Smith is insane*, but rather:

(*) *Over there next to that Buick, there is a* man x dressed like a pirate, jabbering incoherently, brandishing a saber, and x is insane.

Now let us focus on your utterance of "that man is sane". Similar remarks apply. You weren't just seeing Smith; you were seeing *a* man who was wearing a tuxedo, and saying very intelligent things in a very erudite manner. And you saw him as being under a certain chandelier, next to a certain Rothko painting. What we just said about what you is true (give or take a few nuances relating to spatial location) to the person to whom you addressed your comment. So what was communicated by your utterance was:

(**) *Over there, next to that Rothko painting, there is a man x dressed in a tuxedo, saying very intelligent things, and x is sane.*

Obviously (*) and (**) do not contradict each other.

So when we focus on what is communicated by your utterances of "that man is sane" and "that man is insane", we see that you were not guilty of incoherence in virtue of making those statements (even though their literal meanings contradict each other). But consideration of the character-content distinction is not enough: we had to take into account the descriptive nature of sense-perception; we have to be very careful in our description of what it was, exactly, that was seen.

This helps solve a problem that Kaplan himself identified for his own theory. In the late morning, you point up at the sky, at the last celestial body to remain visible. You say "that" very slowly. After a few hours, you say "is identical with". Then, the moment a celestial body becomes visible in the evening sky, you say "that". So you have said "that is that", though very slowly. The referent of the first "that' is identical with the referent of the alter. So, if Kaplan's theory is right, then that utterance encodes a trivial proposition, one of the form: *O is identical with O.* (In fact, it encodes the proposition *Venus=Venus.*) But obviously what was *communicated* was not trivial at all.

How do we deal with this? The character-content distinction isn't much help, since the same character, and the same content, is involved in both cases. Here we need to apply the same method as before. Consider the first utterance of "that". What was the visual perception you had in connection with that utterance? You weren't *just* seeing Venus. You were seeing *a* celestial body that alone occupied the late morning sky. (Of course, you were seeing much more than this: imagine all the data that go into a visual perception: the specificity of it – all the spatial and chromatic content.) So what was pointed out was not just *Venus*; what was pointed out would be given (partially, though not completely) by the expression "occupant of the late morning sky", or some such. And, for exactly similar reasons, the perception you had in connection with your second utterance of "that" was not (just) *Venus*; what was pointed out would be given (partially, though not completely) by the expression "sole occupant of the early evening sky", or some such. So your utterance of "that is identical with that" implicated an existential proposition: *there is some sole occupant x of the late morning sky, and there is some sole occupant y of the early evening sky, and x=y.* So even though what was literally meant by your utterance of "that is identical with that" is simply *Venus=Venus*, what was communicated was a non-trivial, non-analytic existence claim. To sum up, once we take into account the descriptive (or "predicational") nature of sense-perception, along with the fact that utterances have to be *interpreted* in light of such information, it becomes clear what indexical utterances communicate. In connection with this, it becomes clear why utterances that, at the level of literal meaning, are absurd or trivial, may be non-trivial or true (or both) at the level of implicature.

Later we will see how this line of thought applies to Kripke's famed Pierre-paradox and how it rehabilitates an age-old conception of rationality.

§ For the moment, let us adopt a Fregean semantic stance. This involves suppressing our sensitivity to the distinction between type- and token-semantics.

The distinction between "sense" and "reference" is close to, if not identical with, the distinction between "intension" and "extension". The *sense* of "the inventor of bifocals" is the concept (or property) *unique inventor of bifocals.* The referent is the thing to which that concept applies (or which has the property). The sense-reference distinction applies to singular terms.

The intension-extension distinction is the corresponding distinction

for plural terms. Consider the expression "rhenates". The intension is the concept *thing with a kidney* (or, perhaps, the property of having a kidney). The extension is the class of things having kidneys. The intension of "rhenate" is different from the intension of "chordate". But they have the same extension.

We found that, where definite descriptions are concerned, "sense" is type-semantics and "reference" is token semantics. There is no one entity that has both sense and reference. The type has sense. The token has reference. Where plural referring terms are concerned, an exactly similar thing is true of intension and extension. There is no *one* expression that has both intension and extension. The type has intension. The token has extension. It would be incoherent to say that some *one* expression has both intension and extension: that position leads to a contradiction.

Let us start with a more straightforward case of an expression that is alleged to have both extension and intension.

The word "whale", we are told, has both intension and extension. The intension is a property or function (from objects to truth-values). The extension is the class of whales (Shamu, Willy, Moby Dick…)

We don't want to say that the expression-type "whale" denotes *different* classes in different worlds, while having the same semantics. That is perilously close to absurdity. The *extension* of a term is the set of objects that it refers to. So unless we divorce reference from semantic content, it is self-contradictory to say that "whale" has the same semantics in worlds where it has different extensions.

What we must say is: *tokens* of "whales" denote different classes in different worlds; the *type* has an invariant meaning: it pairs off a token of "whale" in a world W with a class C iff C is the class containing all and only whales in W.

The type "whale" has an intension. That intension of "whale" is a function from worlds to classes: a function that assigns truth to a class C in a world w exactly if C comprises all the whales in w and nothing that is not a whale.

Tokens of "whale" have *extensions.* In this world (at this time), it has as its extension the class containing all and only: *Shamu, Willy…*

But there is no one entity that has both intension and extension. Tokens of "whale" have extension; the expression-type "whale" has intension. Nothing has both. If we say otherwise, then we must say that

"whale" can have the same semantics in both w1 and w2, even though it picks out one set of objects in w1 and a different set of objects in w2. So we would end up having to say that "whale" has the same semantics in both worlds, but picks out different things. To avoid self-contradiction, we must then say that what an expression picks out is no part of its semantics. But surely that is absurd – even though a century of Fregeanism has brainwashed us into thinking otherwise. What could be more integral to the semantics of a referring term than the identity of its referent? Frege's answer is – "its sense". But we've seen that Frege's view isn't necessary to explain the facts it is supposed to explain (facts relating to cognitive significance). We've also seen that Frege's view leads to an incoherent conception of reference (reference ends up being quantification). So we can once again embrace our pre-Fregean naiveté. We can once again say:

(*) "Given a referring term, the identity of the referent *is* essential to the semantics of that term."

The difference in cognitive value between a token of "rhenates are warm blooded" and "chordates are warm-blooded" is given an explanation exactly analogous to that previously given in connection with tokens of "the inventor of bifocals snored" and "the first post-maser general snored."

A qualification is in order here. I myself think that, in some cases, tokens of "rhenates" *do* function as quantifiers. Supose you say "rhenates are necessarily chordates". There is no set of individuals $A_1 \ldots A_n$ such that you are saying: $A_1 \ldots A_n$ are necessarily chordates (even though it is probably true that they are necessarily chordates). You are saying: necessarily, for any x, if x is a chordate, then x is a rhenate. So I think that, in some cases, tokens of "whales", "rhenates", and the like, are possibly quantifiers.

Let us sum up. (*) is a reasonable view. If we accept (*), then it becomes contradictory to say that "whale" can vary in extension (referent, basically) while keeping the same semantics. At the same time, it is clear that "whale" has *both* intension *and* extension; and that both are essential. We accommodate all of this by distinguishing types from tokens. The type has intension (a function from worlds to sets); the tokens have extension (the just mentioned sets). Nothing is required to have the same semantics while having different referents. So there is no contradiction; and there is no need to avoid

contradiction by adopting the desperate view that the referent of a referring term is irrelevant to its semantics. The data relating to cognitive significance are easily explained: we need only take into account the differences in *type*_-semantics relating to (say) "rhenates" and "chordates", along with the fact that token-meaning is always grasped by way of knowledge of type-semantics.

§ Carnap (1947) said that the intension of a "sentence" (he didn't distinguish tokens from types) is a proposition, and that the extension of a sentence is a truth-value.

This view is generally seen as either true or, if not quite true, then very reasonable: a reasonable extension of the term "extension".

Carnap's view, I am afraid, doesn't deserve quite such a warm response. We found that the *intension* of "whale" is a property of the corresponding type, and that the *extension* is a property of tokens (in a given world). Types have intention; tokens have extension.

By parity of reasoning, sentence-types have *intension*; sentence-tokens have extension.

The intension of a sentence is some kind of function. But it cannot be a function from worlds to truth-values, for the reasons we've seen. The type "I am tired' does not bear any proposition; so it cannot have a proposition for its intension. The same must be true of the type "Socrates was wise". If we assign a different kind of intension to the type "Socrates was wise", we end up committed to the highly arbitrary view that sentences can have entirely different sorts of things for the intensions: some have propositions (functions from worlds to truth-values), while others have functions from contexts of utterance to truth-values.

In so far as the type "I am tired" has a meaning, it is a function from contexts of utterance to propositions. So if anything is its intension, it is that function. If anything is the intension of "Socrates was wise", it is a function from contexts (worlds, times – there may be other parameters) of utterance to propositions.

The *extension* of a token of "I am tired" must be the thing that the *intension* of the corresponding type assigns to that token – just as the *extension* of a token of "whale" is the thing that the intension of the corresponding type assigns to that token. So the extension of a token of "I am tired" is the *meaning* of that token. Thus, the extension of a token of

"Socrates was wise" is the thing it means (roughly, the proposition *that Socrates was wise*); it is not, *pace* Carnap, a truth-value.

A *proposition* can be seen as a function from worlds to truth-values. So, at least arguably, the extension of *Socrates is wise* is the truth-value *true* in worlds where he is wise, and the truth-value *false* in worlds where he is not.

But the extension of a token of "I am tired" is not a truth-value. Its extension is the thing which the function associated with the corresponding type assigns to that token. That thing is, very roughly, a proposition. That proposition *in its turn* has a truth-value. But the truth-value is not the thing assigned to that token by the corresponding type. So Carnap's view is simply wrong.

Also, I don't think it is quite correct to say that *propositions* have extensions. Intensions and extensions are things that expressions have. Propositions aren't expressions. Strictly speaking, it is no more accurate to say that *Socrates is wise* has the "extension" *true* for this world as an argument than it is to say that "F(x)=2x" has the "extension" four for two as an argument.

Before he did his work on indexicals, Kaplan[81] once wrote:

[T]he question of the truth value of which a given proposition actually is a concept and the individual of which a given individual concept actually is a concept are…empirical. That is, although a given proposition may actually be a concept of Truth, there are usually other possible states with respect to which it is a concept of Falsehood.

A proposition is indeed a concept. But it is a concept of a world, not of a truth-value. A concept of Socrates is some concept C such that Socrates falls under C. *x is a greatest teacher of Plato* is a concept of Socrates because Socrates falls under that concept. Given that Socrates falls under that concept, we can say that it assigns *truth* to Socrates. A concept is a function from objects to truth-values. *x is a greatest teacher of Plato* is not a concept of the truth-value that it assigns to Socrates; it is a concept of Socrates.

If a proposition is a concept, it is a function that assigns truth-values to objects. If a proposition is such a function, the objects in question are worlds. Indeed, propositions are often analyzed as functions that assigns

truth-values to worlds. Let w be this world. The proposition *Socrates was wise* assigns truth to w. Just as *x is a greatest teacher of Plato* is a concept of Socrates, not of the truth-value that it assigns to Socrates, so *Socrates was wise* is a concept of w, not of the truth-value that it assigns to w. Propositions are concepts of worlds, not of truth-values.

When we make it clear what concepts are – functions from entities to truth-values – it is no longer tenable to say that the semantic contribution of a sentence-token is a truth-value.

§ This brings us to one other reason to say that sentence-tokens refer to *their meanings*, not their truth-values. I don't think this reason is conclusive. But it is suggestive.

The content of the type "inventor of bifocals" is a rule that assigns a referent to its tokens. The content of a given token is the referent. The same is true of "Plato". The only difference is that semantic content of the type "Plato" is a constant function, whereas that associated with "the inventor of bifocals" is a variable function. In some world, tokens of "the inventor of bifocals" refer to Jefferson. In all worlds, tokens of "Plato" refer to Plato. (We are talking about worlds semantically identical with ours.) So when we look at singular terms, we find that sense (intension) is identical with type meaning and referent (extension) is identical with token meaning.

So if we are to apply the sense-reference (intension-extension) apparatus to whole sentences, this is what we find. Intension is a property of sentence-types. Extension is a property of sentence-tokens. The extension of a sentence-token is the thing which the corresponding type assigns to it. So the extension (referent) of a sentence-token is its meaning. The referent of a token of "you" is the thing which the corresponding type assigns to that token. If we wish to find a sentence-level analogue of this, it is clear what we must say: reference is a property of sentence-tokens; sense is a property of sentence-types; the referent of a token of a sentence is the meaning assigned to it by the corresponding type; the sense of the corresponding type is a rule that assigns a meaning to that token.

Here some words of clarification are in order. As we said earlier, we must distinguish between forceless sentence-tokens and forced sentence-tokens. I believe that forceless sentence tokens really do refer to propositions, It isn't that the relation of such expressions to propositions is some kind of *analogue* of the relation that holds between tokens of "Socrates" and

Socrates. We are dealing with an *identity*, not an analogy. In a token of "if snow is white, then something is white", the occurrence of "snow is white" really does refer to the corresponding proposition.

But as we discussed earlier, I think that, where forced sentence-tokens are concerned, we *are* only dealing with an analogy: such sentence-tokens do not refer to anything. (In any case, this is what I tend to believe. Later I will state some reservations.) Consider a forced token of snow is white. For the sake of argument, suppose that it refers to some object O. Now consider an expression that *uncontroversially* refers to O, for example *O* or (supposing that O alone has phi) *the unique phi*. The relation between that token of "snow is white" to O will be different from the relation that holds between *O* and O. And I don't think that this difference in the two relations is merely grammatical.

I don't think that forced sentence-tokens *actually* refer to anything. When we consider any paradigm case of a referring term, we find that it can occur as a proper part of a sentence-token. Tokens of "Socrates" can occur as proper parts of sentence-tokens. Indeed, if they do not so occur, what results is ill-formed. There are cases where saying "Socrates" by itself constitutes a well-formed utterance. (This would be the case if you said "Socrates" in response to "who was the greatest philosopher of antiquity?") But in any such case, that utterance is elliptical for a whole sentence ("Socrates is the greatest philosopher of antiquity").

A forced sentence-token cannot occur as a proper part of another sentence. This suggests that such tokens are in a very different category from expressions that we know to refer.

This suggestion is reinforced when we ask what it is for an expression to refer. This is what we found before. Tokens of E refer to O exactly if, in virtue of having the form "…E…" a sentence encodes a proposition that is *about* O. (Remember that, in our judgment, "Sally met a man" is not really *about* Fred, even if Fred is the man whom Sally met.) So an expression refers to a thing if sentence-tokens containing that expression are about that thing. And E is an expression that *on the whole* refers to O iff, in virtue of having the form "…E…", a sentence-token encodes a proposition that has O as a constituent and there is no O* (distinct from O) such that, in virtue of having that form, a sentence-token encodes a proposition that has O* as a constituent.

Given this, let us talk about forced sentence-tokens. Let t be a forced

sentence-token of "snow is white"…

Summary

All sentences of natural language contain an indexical component, and are thus in the same category as the type "I am tired". Therefore, the meaning or intension of a sentence-*type* is (approximately) a function from contexts of utterance to propositions. (We will see that this is not *quite* accurate.) The value of a sentence-token is (approximately) a proposition. So sentence-tokens have *propositions* for their referents or extensions. They don't have anything for their *intensions.* Expression-types have intensions; expression-tokens have extensions. Sentence-tokens do not refer to truth-values, or have such things for their extensions. Once two-dimensionalism is accepted, it becomes impossible to say that truth-values are the semantic-values, even in part, of either sentence-tokens or sentence-types. It also becomes impossible to deny that propositions are, at least in part, the semantic values of sentence-tokens.

The Julius-problem

Consideration of a puzzle due to Kripke verifies our analysis. We will consider Kripke's puzzle in a form given to it by Gareth Evans.

Let us *define* "Julius" thus:

(J) "Julius" is a name of that person, whosoever it was, that uniquely invented the zipper.

So if Smith invented the zipper, then "Julius" refers to Smith. If Jones invented the zipper, then "Julius" refers to Jones. Here is the puzzle.

(Z) "Julius invented the zipper".

I know *a priori*, says Kripke, that (Z) is true. But (Z) is a *contingent* statement. Whoever Julius is, it is surely a contingent that he invented the zipper. After all, that person might *not* have invented the zipper; he might have died at the age of five. So, says Kripke, I have *a priori* knowledge of a contingent matter of fact. This is a puzzle; in any case, it is quite a shock.

But there is no real puzzle here. We have to distinguish sentences from propositions. Kripke and Evans both use the term "statement", which is ambiguous between propositions, on the one hand, and sentence-tokens, on the other.

For the sake of discussion, suppose Smith invented the zipper. In that case, (Z) semantically encodes the proposition:

(S) Smith invented the zipper.

Nobody knows a priori that Smith invented the zipper. I could know Smith for forty years and not know that he invented the zipper. It is also plain that (S) is contingent.

Given *only* my knowledge of English and, in particular, of the semantic rule for "Julius", I know that (Z) is true. Knowledge of nothing *other* than the semantic rules associated with its constituents is needed to know that (Z) expresses a true proposition. In this respect, (Z) is unusual among sentences. Most sentences that express contingent, a posteriori propositions do not have this property.

But my knowledge of the semantic rules for (Z) is quite definitely *a posteriori.* My knowledge that "zipper" means *zipper*, that "invent" means *invent* is a posteriori.

I *don't* know *a priori* that Smith invented the zipper. If I know that at all, I know it a posteriori. I *don't* know *a priori* what the semantic rules are for (Z). If I know that at all, I know it a posteriori. Nothing here is known a priori. What we have here is *not* a case of a priori knowledge of a contingent truth.

Nonetheless there is a still a puzzle. Given *only* a knowledge of the semantic rules for (Z), we seem both to understand it *and* to know its truth-value. To understand a sentence is, presumably, to know what proposition it encodes. To know the truth-value of a sentence is to know the truth-value of the proposition it encodes. So a knowledge of semantic rules alone enables one to know a matter of contingent, non-linguistic fact. "With the stroke of a pen" we can figure out who invented the zipper. By the same sleight of hand, we can figure out who will cure AIDS. Let

(Y) "Yablo" name that person, whoever it turns out to be, who discovers the cure for AIDS.

The sentence

(CA) "Yablo will cure AIDS"

expresses a contingent truth. But given only a knowledge of semantic rules, I both understand, and know the truth-value, of (CA). Thus a knowledge of semantics alone enables me to know who will accomplish this great feat. But, as Evans said, one cannot generate such knowledge with a "mere stroke of the pen".

Given what we said about definite descriptions, it is perfectly clear how to deal with this puzzle. We need only be clear on (i) the distinction between *type*-meaning and *token*-meaning; and (ii) what it is for an expression to *refer* to an object.

To say that "Julius" *refers to* that person, whoever it should turn out to be, who invented the zipper is to say this:

(JR) *If* somebody x uniquely invented the zipper, then a token of "… Julius…" encodes the proposition:…*x*…

If there is no such person, then such a token is abortive: it encodes no proposition.

Thus, the semantic rule for

(Z) "Julius invented the zipper",

is this:

If somebody x invented the zipper, then a token of *t* of (Z) means:

x invented the zipper,

If there is no such person, then *t* encodes no proposition.

What I understand is the meaning of the sentence-*type*. Given some *particular* token *t** of (Z), to understand that token, I would have to know *which* exact person invented the zipper. To understand *t** is to know, specifically, *which* proposition to associate with it: it is to know whether it

means *Smith invented the zipper* or *Jones invented the zipper*. If I do not have this knowledge, then I don't understand that *token.* But I can lack such knowledge and have a perfect understanding of the corresponding sentence-*type*.

I can*not* know the truth-value of a *token* of (Z) without knowing who invented the zipper. In fact, without such knowledge, I cannot even assign meaning to (Z). I *can* understand the meaning of the corresponding *type* without such knowledge. Given *only* a knowledge of semantic rules, I know that, *if* there is an x such that x uniquely invented the zipper, then a token of that type means: *x invented the zipper*. So when I hear a token *t* of (Z), I know that *if* there is an x such that x uniquely invented the zipper, then *t* means: *x invented the zipper*. So any such token will be replete with cognitive significance. But I still will have no idea what its literal meaning is.

Remember that *sentence-types* are not true or false. Sentence-types encode, not propositions, but functions that assign propositions to sentence-*tokens.* So, contrary to what Evans and Kripke said, given *any* expression that encodes a contingent proposition, one cannot, purely on the basis of linguistic or conceptual knowledge, figure out what its truth-value is. In any case, nothing they have said requires us to reject this view. For the truth-value of a *token* of (Z) can no more be known, wholly on the basis of logical and semantic knowledge, than can the truth-value of a token of "Smith invented the zipper." The sentence-*type* has no truth-value at all. So under no circumstances does purely semantic knowledge yield knowledge of the truth-value of *anything* non-semantic. Puzzle dissolved.

Kripke thought himself to have shown that there is *a priori* knowledge of contingent facts. We've seen that Kripke has shown no such thing; and we are not alone in concluding this. But it is generally accepted that Kripke showed there to be analytic, a posteriori *sentences.*

There is *some* truth in this. But to make it clear what that truth is, we need to very clear about the difference between sentence-types and sentence-tokens.

Some sentences are analytic, and some are not.

(T) "triangles have three sides".

is *analytic.* What does this mean? The pre-Kripke answer was:

(T) encodes the proposition: triangles have three sides. Given only a grasp of that proposition, one can know whether it is true or not. In general, a sentence is analytic iff it encodes a proposition P such that, given only a grasp of P, one has enough information to decide the truth-value of P. A sentence is analytic iff it encodes a "conceptually" true proposition.

But given what Kripke said, that is not a correct analysis. Perhaps it provides a *sufficient* condition for analyticity, but not a necessary one. After all, (Z) is plausibly regarded as "analytic". But (Z) encodes a proposition that is *not* "conceptually true". *Smith invented the zipper* is as empirical as can be.

Given what Kripke said, the concept of analyticity must be re-analyzed. The re-analysis is not hard to produce. But some preliminary points are needed to forestall confusions. First of all, semantic knowledge is always empirical (a posteriori). The only way to know that "Socrates" refers to Socrates is through sense-experience. "Socrates" could refer to anything or to nothing.

Also, a sentence is always a sentence *of* this or that language. So "triangles have three sides" is analytic *in English.*

Given this, there is an obvious Kripke-friendly analysis of analyticity:

(EM) Given a sentence S belonging to a language L, S is analytic exactly if the only empirical knowledge needed to know whether S is true is knowledge of the semantic rules of L.

I am actually going to amend (EM). But first let me give the *raison d'être* for it. Consider

(*) "12+50=62".

By itself, knowledge of the *semantics* of (*) will not enable one to decide its truth-value. One has to know whether 12 and 50 make 62. That knowledge is mathematical, not semantic. But the only *empirical* knowledge one needs to know the truth of (*) is the knowledge that "12" denotes 12, "=" denotes equality, and so forth: it is entirely linguistic. So (*) is analytic because no *empirical* knowledge, other than knowledge of the semantic rules that give it meaning, is needed to decide its truth-value.

So (Z) is analytic because, given *only* a knowledge of the semantic rules that give it meaning, one has all the *empirical* knowledge needed to know that it is true. Of course, other kinds of knowledge – a certain rudimentary knowledge of logic – is needed to supplement that empirical knowledge, if one is to know that (Z) is true. But any such non-semantic knowledge will be non-empirical.

This analysis fits the data. Consider the sentence:

(SOC) "Socrates is named ‘Socrates’".

The proposition that (SOC) encodes is entirely *a posteriori.* Socrates might not have been named "Socrates". He might not have been named anything. But (SOC) is obviously analytic. For given *only* a knowledge of the semantic rules that give it meaning, one has all the *empirical* data one needs to decide its truth-value. Among those rules is: for any x, "Socrates" refers to x iff x is identical with Socrates. Once one knows that rule, and a few others (e.g. "names" denotes a certain relation), one has all the *empirical* knowledge one needs to decide the truth-value of (SOC). For exactly similar reasons,

(SW) ‘there is a language L such that, in L, "snow is white" means: *snow is white*",

is analytic, even though the proposition it encodes is a posteriori. There didn’t have to be *any* language of the kind described. "snow is white" might have been meaningless.

Kaplan produced many examples of analytic sentences that encode a posteriori propositions, for example:

(DS) "dthat [the tallest spy] is a spy".

It is easy to verify that (EM) applies to (DS).

There is additional confirmation for (EM):

(SU) " ‘schnee ist weiss’ means: *snow is white*"

is *not* analytic. This is consistent with our analysis. Our analysis says: if S belongs to L, then a knowledge of the semantic rules of *L* is the

only empirical knowledge needed to decide the truth-value of S. (SU) does not satisfy this requirement. (SU) is a sentence of *English.* It contains a German expression; but, clearly, it is itself an *English*, not a German, sentence. But one must have empirical knowledge of something *other* than English-semantic rules to decide as to its truth-value: one must have knowledge of *German* semantic rules. So our analysis correctly predicts that (SU) is non-analytic.

Nonetheless, (EM) is subject to a modification; and this modification undercuts the importance of Kripke's discovery. For reasons we have seen, it is meaningless to say that a "sentence" is analytic. There are sentence-tokens, and sentence-types. There is no such thing as a "sentence" *simpliciter.* It is meaningless to say that (Z) is analytic. One must say either that *tokens* of it are analytic or that the *type* is analytic.

Obviously the *type* is not analytic. For an expression to be analytic, it must be true or false. A sentence-*type* is never true or false. There *is* a sense in which a token of (Z) is analytic. Given any such token, the only *empirical* knowledge one needs to decide its truth-value is knowledge of the semantic rules of the language to which it belongs. So the right definition of "analyticity" is this:

(EM_T) Let S be a sentence-type belonging to L. A token t of S is analytic exactly if the only empirical knowledge needed to decide the truth-value of t is knowledge of the semantic rules for L.

Given a token *t* of (Z) or (SOC) or (SU) or (DS), the only empirical knowledge one needs to decide its truth-value is knowledge of the semantics of the language to which they belong. (In this case, that language is English, plus a few extra semantic rules invented by philosophers.)

Chapter 11 Are all terms referring terms?

I have argued that reference is semantic contribution. There is no difference between referring to a thing and semantically contributing it. We referred to this principle as SC. If SC is right, then it follows immediately that sentence-tokens refer to their meanings.

But I doubt our arguments for SC have addressed everyone's misgivings about it. The main misgiving, I think, is this:

Maybe you're right to say that tokens of "Plato" refer to Plato exactly if Plato is semantically contributed by such tokens. So in some cases meaning collapses into reference. But SC says that meaning always collapses into reference, and this is plainly false. Consider the word "and". This word has meaning. But surely it doesn't refer to anything. It has meaning by virtue of having some kind of syntactic or formal function. It does not have meaning by virtue of picking out some object.

You have a Neanderthal conception of meaning: to mean is to denote. This is exactly the position that Wittgenstein refuted in both the Tractatus and in his later works. In the Tractatus, he refuted it in connection with connective terms like "and" and "or". In his later work, he refuted it in connection with lexical items like "Nixon" and "Socrates".

Semantic content is not *always* identical with reference. Expression-types, I have argued, do not refer to anything; but they have semantic contents: functions of some kind.

But with that qualification, I do have exactly the conception described. Where expression-tokens are concerned, to mean is to denote. When we say that some expressions have a purely "formal" or "syntactical" function, we are really saying that they denote special kinds of things: second-order functions, to be precise.

Consider the word "and". The standard view about it is this:

You don't define "and" by pointing to some object that it denotes. You define it contextually. You say what it means by saying what is meant by whole sentences containing it.

The idea seems to be that if "and" referred to something, you could define it non-contextually; you could simply indicate what it denoted. Since it must be defined contextually, it doesn't refer to anything.

This view is sheer folly. For "Smith" to pick out Smith is *precisely* for "Smith" to admit of a certain kind of contextual definition. It is precisely

for "…Smith…" to mean:…*Smith*…

Of course, "Smith" can be defined ostensively, whereas "and" cannot. But ostensive definition is a form of contextual definition. When you say:

(*) *That* person is named "Smith"

that is just a condensed way of giving a contextual definition. You are saying:

(**) Consider *that* person over there. Let O be that entity. "…Smith…" means:…*O*…

This can be shown by analogues of arguments already given. Suppose you point to some object O and you say:

That person [pointing at O] is named "Smith". But if you want to make a statement about him, you cannot use the word "Smith" to do so. "Smith smokes" doesn't mean that he [pointing at O] smokes; it means that some other person smokes. In general "…Smith…" doesn't mean that he [pointing at O] has:…x…

That would be absurd. You started out by giving an ostensive definition. But you then stripped that definition of any force it initially had.[82] So if you point to O and say "*that* is Smith", a necessary condition for your definition to have any force is that "…Smith…" mean:….*O*…

That is also a sufficient condition. Suppose you said:

Sentences of the form "…Smith…" mean that that person [pointing to O] has:…x…But "Smith" does not name that person.

That would be absurd, for now familiar reasons.

When you give an ostensive definition, you are really giving a condensed contextual definition.

Of course, pointing cannot be involved in a definition of "and" or "or". But that has nothing to do with semantics; that has to do with metaphysics and epistemology. We cannot point to abstract objects or to

objects existing in the future. But we can refer to such things.

Perhaps we must define "and" contextually. But this is of no semantic consequence. "Smith", no less than "and", is always defined contextually.

Strictly speaking, we can no more point to the meaning of "Smith" than we can point to the meaning of "and". When we say that Smith is the meaning of "Smith", we are really making a statement about a class, an infinitely large class, of expressions. Smith's being referred to by "Smith" – Smith's being the meaning of "Smith" – is really identical with the fact that sentences of the form "…Smith..." mean:…*Smith*…We can more point to *that* fact than we can point to the meaning of "and".

The other side of the coin is that any contextual definition can be turned into a non-contextual or denotative definition. Frege rightly said that "some man" does not denote some man. He concluded that it must be defined contextually:

(i) "…some man…" is true exactly if: *for some x, x is a man and… x…*

This is what people typically say about "and" and case-markers, and other so-called non-denoting expressions. This is taken to imply that such expressions do not denote anything. But it is easy to convert any contextual definition into a denotative definition. For example, (i) is equivalent to:

(ii) For any concept C, *C(some man)* is true exactly if for some x, x is a man and C(x).

(ii) is equivalent to:

(iii) for any concept C, "some man" assigns truth to C exactly if for some x, x is a man and Cx.

(iii) is equivalent to:

(iv) "some man" denotes a function that assigns truth to a concept C exactly if, for some x, x is a man and Cx.

What we just said about "some man" can be done with *any* expression. Any contextual definition can be reduced to a denotative definition. If an expression can be defined contextually, it denotes something.

In effect, we've already seen this. Connectives like "after", "and", and "because" denote functions. The same thing is true of tense- and case-markers.

Granted, there is a difference between expressions like "Plato" and expressions like "and". But the difference is not that the former denote, while the latter do not. The difference lies in *what* they denote. The former denote individuals. The latter denote functions.

Actually, we will see some reason to believe that the difference is even *more* tenuous than this suggests. Arguably, tokens of "Plato" *do* denote functions, no less than "and". The difference between tokens of "Plato" and tokens of "and" lies in what kind of functions they denote.

The Augustinian conception of language, ridiculed by Wittgenstein, turns out to be quite right.

Since reference reduces to semantic contribution, it follows that sentence-tokens refer to what they semantically contribute; they refer to what they mean. In any case, if one is to escape this conclusion, one must show that SC is wrong. But this does not appear to be possible.

Is it a truism that meaning is compositional?

Many (most famously Frege) have maintained that the meaning of a sentence is a function of the meanings of its parts. Linguistic meaning is "compositional". This is known as the "thesis of compositionality" or, simply, "compositionality".

Compositionality has been accepted by most, but not all. To my knowledge, it has always been regarded as a substantive thesis. I think that compositionality is correct. But I also think it is a truism, in the same category as "we *can* travel in time: at the rate of one second per second".

We have argued that, for (tokens of) "Socrates" to refer to Socrates just *is* for sentences of the form "…Socrates…" to have a certain meaning. If they *don't* have that meaning, then "Socrates" didn't refer to Socrates to begin with. There is no way that a sentence-token could not be a function of the meanings of its parts. An expression means such and such exactly if sentences containing it have certain meanings. So whatever meanings those

sentences have, that fixes what the meaning of that expression is. Under no circumstance can we meaningfully, let alone truly, say that E means O but that "…E…" does not mean:…*O*…Compositionality is not a thesis about subsentential meaning; it merely identifies what subsentential meaning is.

Is the term "direct reference" a pleonasm?

As we noted, a token of "you", addressed to (say) Benjamin Franklin, semantically contributes Franklin himself, and not a concept that applies to him. This is also true of tokens of "Benjamin Franklin" and possibly of "the inventor of bifocals".

The doctrine that some terms contribute *individuals*, as opposed to concepts applying thereto, is called "direct reference theory".

But really "direct reference theory" should just be called "reference theory". If tokens of "the inventor of bifocals" contribute the concept *inventor of bifocals*, as opposed to Franklin, then they would not refer to Franklin at all; they would be quantifiers, not expressions that referred to Franklin. Reference is always "direct".

The debate between friends and foes of "direct reference theory" is really a debate as to whether there is such a thing as reference; it is not a debate as to the nature of reference.

Gödel's Slingshot

According to Gödel[83], the Slingshot shows *either* that there is one fact *or* that Russell's theory of descriptions is correct. Given this ugly choice, Gödel rightly infers that Russell's theory is correct.

But there is no need to make that choice. For Gödel's Slingshot involves the muddle we've been discussing.

Let "Fa" and "Gb" be any two true, non-analytic sentences. Assume that a is not identical with b. Finally, assume that the meaning of a referring term is the thing it refers to.

1. Fa. *premise*
2. a is the unique thing x such that [Fx and x=a]. *co-refers with 1 by LL*

3. a is the unique thing x such that [a is not b and x=a]. *co-refers with 1 by LL and CR*

4. a is not identical with b. *co-refers with 3 by LL*

5. b is the unique thing y such that [y is not identical with a and y=b]. *co-refers with 4 by LL*

6. b is the unique thing y such that [Gy and y=b]. *co-refers with 5 by CR*

7. Gb *co-refers with 6 by LL.*

If cogent, Gödel's argument shows that *either* definite descriptions must parse out, in the way described by Russell's theory, *or* that any two true sentences have the same meaning. If Russell's theory is right, then CR cannot be used to generate the needed transitions. Obviously not all sentences have the same meaning. So Russell's theory must be right.

There are two major problems with Gödel's argument.

First, that argument relies heavily on the idea that logically equivalent sentences co-refer. This assumption is unwarranted – it appears warranted only when S-equivalence and L-equivalence are conflated -- and it is probably false. So Gödel's argument fails.

Let us discuss the second problem with Gödel's theory. This one is identified by Barwise and Perry (1983). Tokens of definite descriptions are either quantifiers or they are terms that refer to individuals. Let us consider each case.

First let us suppose that tokens of definite descriptions refer to individuals. Remember what we established earlier: if two terms co-refer, they make exactly the same semantic contribution. The following two expressions co-refer:

(i) "a"
(i) "the unique thing x such that [Fx and x=a]"
(iii) "the unique thing x such that [a is not b and x=a]."

So they both contribute a and a alone.

The following three expressions all co-refer.

(iv) "b"
(v) "the unique thing y such that [y is not identical with a and y=b]".

(vi) "the unique thing y such that [Gy and y=b]."

So they all contribute b and b alone.
Thus Gödel's argument becomes:

1. Fa. *premise*
2. a is a. *co-refers with 1 by LL*
3. a is a. *co-refers with 1 by LL and CR*
4. a is not identical with b. *co-refers with 3 by LL*
5. b is b. *co-refers with 4 by LL*
6. b is b. *co-refers with 5 by CR*
7. Gb *co-refers with 6 by LL.*

But this is an obvious failure. "Fa" is not logically equivalent with "a is a". (Remember that "Fa" is non-analytic.) So even if logically equivalent sentences *did* co-refer, (1) and (2) would not co-refer. For the same reason, neither would (6) and (7). So the argument is a failure if definite descriptions refer to individuals.

It is a failure if they don't refer to individuals. If they don't refer to individuals, they are quantifiers. If they are quantifiers, then either they don't refer to anything or they refer to functions. If they don't refer to anything, then (CR) no longer permits the steps from (2) to (3) or from (5) to (6).

If definite descriptions refer to functions, then CR *still* cannot be used to make those steps. If they refer to functions, then "the unique thing x such that [Fx and x=a]" refers to a very different function from "the unique thing x such that [a is not b and x=a]." There are possible worlds (this is not one of them) where English is spoken where those functions assign different truth-values to the same entity.

On any reading of definite descriptions, Gödel's argument fails.

There is yet another reason why Gödel's argument fails. Replace each occurrence of "the" in that argument with "ze", as defined earlier. Let Gödel_Z be the resulting argument. Gödel_Z is an exact analogue of Gödel's original argument. Nothing can *prove* that "ze phi" is not a referring term; for, by our stipulation, it is just that. But *if* Gödel_Z is cogent, then it "proves" that "ze phi" doesn't refer. If Gödel_Z is cogent, it isn't cogent; therefore it isn't cogent. Therefore, the same is true of Gödel's original argument.[84]

Chapter 12 The semantics of non-indicatives

For reasons just discussed, the meaning of a sentence-token is *in part* a proposition. So part of what a sentence-token refers to is a proposition. But that is not *all* it refers to. A correct semantics must say *exactly* what a sentence-token refers to.

Two sentence-tokens can encode exactly the same proposition but differ profoundly in their literal meanings.

(i) "Plato snores"

is true or false.

(ii) "Does Plato snore?"

is neither true nor false.

But both encode exactly the same proposition: *that Plato snores.*

Further, tokens of

(iii) "Plato, snore!"

and

(iv) "that Plato snores"

encode that very same proposition. But no two of them have quite the same literal meaning.

(iv) isn't a sentence-token at all. It doesn't assert, order, or ask anything.

A correct semantics must account for these differences among (i)-(iv).

We've identified reference with semantic contribution. What an

expression-token refers to is nothing other than what it semantically contributes to sentences in which it occurs.

Tokens of (i) are assertions; they have indicative force. Tokens of (ii) are commands; they have imperative force. Tokens of (iii) are questions; they have interrogative force. Tokens of (iv) *that Plato snores* don't have any of these properties.

So tokens of (i)-(iv) all contribute *different* things. If our analysis of reference is correct, they cannot all refer to the same thing.

At the same time, there is not the slightest doubt that at least *part of* what is semantically contributed by tokens of (i)-(iv) is the proposition: *that Plato snores.* So *part of* what they refer to is that proposition.

Here is how I propose to make sense of all this.

Part of what tokens of (i)-(iv) refer to is the proposition: *that Plato snores.* The reason is that all those tokens actually *contain* the expression: "that Plato snores". (i)-(iii) contain it in a phonetically disguised form. Suppose somebody asks you "who was the greatest philosopher of antiquity?", and you say "Socrates". Your answer contains a phonetically disguised: "was the greatest philosopher of antiquity." Similarly, (i)-(iii) contain a phonetically distorted "that Plato snores".

But (i)-(iii) do not *just* contain that expression. They contain extra semantic material. In each case, that material denotes a function. That function is what gives indicative force to tokens of (i), imperative force to tokens of (ii), and so on. We will work out the details shortly.

A correct semantics must do justice to the fact that "Plato snores", "does Plato snore?", and "Plato, snore!" are all members of the same basic semantic category. They are all sentences. There are important differences between them. This puts them in different *sub*-categories. But there is some one non-disjunctive category that comprises all and only those things we call "sentences". A correct semantics must find the property had by all and only the things in that category.

A sentence-token, I will argue, is an expression that denotes a very special kind of function.

I want to forestall a couple of mistaken views that might be had in connection with the matters just described. Then I will give my own analysis.

Differences in force are not differences in propositional content

How are we to explain the fact that

(i) "Plato snores"

is true or false, while

(ii) "that Plato snores"

is neither? How are we to explain the fact that they have *different* forces?

There is one obvious answer. One might say that they actually *do* contribute different propositions, and that this is why they differ in force. One might say that, in general, differences in force *do* reduce to differences in propositional content. The idea would be this:

(i) contributes the proposition: it is true that Plato snores, while (ii) merely contributes the proposition: that Plato snores. This is why (ii) is neither true nor false, while (i) is true or false. A difference in force does reduce to difference in the proposition contributed.

But that is absurd. Even if we grant that "Plato snores" contributes the proposition *it is true that Plato snores*, as opposed to (merely) *that Plato snores*, that has nothing to do with the difference in *force* between "Plato snores" and "that Plato snores". For the expression "that it is true that Plato snores" is neither true nor false. But, uncontroversially, the semantic contribution of that expression is the proposition: *that it is true that Plato snores.* Differences in *force* are not reducible to differences in propositional content.

Analogous remarks apply to

(iii) "does Plato snore?"

and

(iv) "Plato, snore!".

What we just said (mutatis mutandis) shows that these sentence-

tokens do *not* differ from "that Plato snores" in propositional content; they differ in some other, completely different respect.

Force not merely a way of signaling a propositional attitude

Here is a view that has elements of truth, but is wrong overall:

A force-operator is a conventionalized way of expressing an attitude towards a proposition. "snow is white" encodes the proposition that snow is white. At some level, "snow is white" contains a force-operator. The presence of that force-operator is the conventional way of expressing one's belief in that proposition.

An exactly similar line can be taken for questions and imperatives. "Is snow white?" can be seen as comprising a force-operator: one giving interrogative force. The presence of that operator is a conventional way of expressing one's wish to know whether a certain proposition is true. The proposition, of course, is: that snow is white.

There is truth in this position, but it is insufficient. Force-operators *are* ways of indicating propositional attitudes. But that is not all they are. Consider the sentence:

(v) "I wish to know whether Plato snores."

Surely that sentence is a conventionalized way of expressing a desire to know whether the proposition *that Plato snores* is true. But an utterance of (v) is semantically very different from an utterance of

(iii) "does Plato snore?"

Semantically, (v) either true or false. Semantically, (iii) is neither. So what makes (iii) be a question is not merely that it is a conventionalized way of expressing one's desire to know whether a certain proposition is true. That is probably part of the story, but not all of it.

Another example might help. Consider:

(vi) "I believe that Plato snores"

A token of (vi) is a conventionalized way of expressing one's belief in the proposition *that Plato snores*. But a token of (iv) is semantically very different from a token of:

(i) "Plato snores".

A token of (vi) encodes a proposition that is true just in case a certain individual has a certain belief. A token of (i) encodes a proposition that has nothing to do with what anyone believes. (i) could be true in a world where nobody had any wishes. (vi) could not.

So it is not enough to say that force-operators are conventionalized ways of expressing propositional-attitudes.

Force is a matter of literal meaning

Another view I want to head off is this:

You say that there are semantic differences among tokens of "Plato snores", "that Plato snores" "does Plato snore?", and so on. I think this is false. Obviously they are all associated with the same proposition and the same truth-value. Obviously they all contribute the same proposition. The meaning of a sentence-token obviously just is the proposition it encodes. This is a truism. The meaning of a sentence-token just is its semantic content. This is a truism. So it follows from truisms that there are not semantic differences among "Plato snores", "does Plato snore?", and so on. Perhaps there are pragmatic differences: differences in what they do (perlocutionary differences). But there are no differences in what they mean. You've made a case that sentence-tokens semantically contribute propositions, and that reference reduces to semantic contribution. So instead of fiddling about with irrelevant pragmatic notions like sentence-force, you should just end your story right here.

The problem with this proposal is that "Plato snores" quite plainly

does differ in literal meaning from "does Plato snore?" and "that Plato snores". Suppose you want to assert that Plato snores. If you say "does Plato snore?" or "Plato, snore!", you are obviously guilty of linguistic incompetence: you obviously don't know quite what is meant by the expressions you are using. So we are not dealing with mere pragmatic or perlocutionary differences.[85]

Whenever an expression has force, that is because it comprises a force-operator. Force-operators, I will argue, are just like other expressions: tokens of them refer to functions of a certain kind. But they are not the functions one might think.

The meaning of a sentence-token

By a "speech act" I mean the tokening – the uttering or writing -- of a sentence. If you say "did Plato snore?" or "Plato snores", you are performing a speech-act.

Many different things can be achieved through speech acts. Money can be made, foes can be wounded, lovers can be wooed. Speech-acts bring us many kinds of success.

But very few such successes fall within the province of semantics. Semantics is concerned only with literal meaning. Suppose I say

(i) "Plato snores".

I can use these words to accomplish many different things - to amuse, threaten, or deceive. But none of this has any relevance to the *semantics* of my words.

But there is *a* kind of success associated with such an utterance that does fall within the province of semantics. It is not part of the *semantics* of an utterance of "Plato snores" that it must amuse or deceive. But it *is* part of the semantics of an utterance of (vi) that it is supposed to be *true.* It *is* part of the semantics of such an utterance that it is a failure if false, and a success if true.

Everyone who speaks English knows as much. Anyone who speaks English knows that an utterance of "Plato snores" is supposed to be true and, thus, that it is a success only if it is true.

In general, there is exactly one kind of success that, at the level of semantics, indicative sentence-tokens are *supposed* to have. They are

supposed to be *true.*

This doesn’t mean that people always aspire to tell the truth when they use indicative sentences. But it does mean that *semantically* they are failing if the indicative sentence-tokens they produce are false.

Under what circumstances is an utterance of "Plato snores" true? It is true exactly if a certain proposition is true. That proposition is: *that Plato snores*. Of course, *that Plato snores* is exactly the proposition that is semantically encoded in such utterances.

Let us take stock. At the level of semantics, there is exactly one kind of success that an utterance of "Plato snores" is supposed to have: it is supposed to be true. Such an utterance is true exactly if the proposition semantically encoded in it (*that Plato snores*) is true. So the semantics of an utterance of "Plato snores" assigns it the property of success exactly if the proposition encoded in that utterance is true. Thus, the *semantic content* of such utterances assigns success to such utterances exactly if the propositions encoded in them are true.

Therefore the semantic content of an utterance "Plato snores" is a function that assigns the property of success to that very utterance exactly if the proposition *that Plato snores* is true.

This is easily generalized. Let *t* be any indicative sentence token. *Part* of what *t* will encode is some proposition P. *t*’s semantic content is a function that assigns the property of success to *t* exactly if P is true.

Force-operators

Utterances of "that Plato snores" are neither true nor false. This is because they do not have assertoric force. In fact, they don’t have any kind of force.

Presumably such utterances lack any constituent that give them sentential force. They do not comprise force-operators.

Utterances of (i) "Plato snores" are true or false because they *do* have assertoric force. So they must comprise some constituent that gives them such force. They must comprise force-operators.

Force-operators cannot be identified with any isolable part of the phonetics of such utterances. This may make people leery of supposing that they comprise such operators. But this is extremely shallow reasoning. Sentences are replete with constituents that have no phonetic representation

or a highly distorted phonetic representation.

We will revisit this shortly. For now, let us operate on the reasonable assumption that, semantically if not phonetically, utterances of "Plato snores" *do* comprise force-operators.

What exactly is it that is given assertoric force?

(a) "Plato snores".
(b) "Does Plato snore?"
(c) "Plato, snore!"
(d) "that Plato snores".

Uncontroversially, *part* of what is semantically encoded in any token of (a)-(c) is the proposition *that Plato snores.*

Given that (a) semantically encodes (among other things) *that Plato snores*, it is not unreasonable to suppose that, semantically if not phonetically, it contains an expression referring to that proposition.

The same is true of (b) and (c). Semantically, if not phonetically, they comprise an expression referring to: *that Plato snores.*

Propositions are not given force. The proposition *that Plato snores* cannot be given interrogative, assertoric, or any other force. Expressions are given force.

So presumably what is going is this. In (a), an expression referring to the proposition *that Plato snores* is given assertoric force. In (b), such an expression is given interrogative force. In (c) such an expression is given imperative force.

Further, given that (a) assigns assertoric *force* to the expression referring to the proposition just mentioned, it is not unreasonable to suppose that, semantically if not phonetically, it contains a force-operator.

So the right way to look at (a) is as a phonetic distortion of:

(a_F) "A(that Plato snores)"

where the "A" is an operator that gives assertoric force to the proposition falling within its scope.

Right now, let us focus on (a) and (a_F). We will return to (b)-(d) in a moment.

Any token of (a_F) is (inter alia) a token of "that Plato snores". That is

why any token of (a_F) semantically encodes: *that Plato snores.*

But any such token also includes a force-operator. Obviously that force-operator gives assertoric force to the occurrence of "that Plato snores" in its scope. But what exactly does this mean? What is it to give assertoric force to an expression?

We've already considered some erroneous answers to this question. Let us now give a correct answer.

(a_F) "A(that Plato snores)".

Let *t* be an arbitrary token of (a_F). Of course, *t* is (inter alia) a token of "that Plato smokes". Any token of "that Plato snores" denotes the proposition *that Plato snores.* The "A" in a given token of (a_F) denotes a function that ascribes success to *t* exactly if the proposition just mentioned is true.

We must distinguish the sentence-*type* (a_F) from tokens of that type. The semantic content of that *type* is a function that, in its turn, assigns a certain kind of function to each token of that type. The semantic content of the type (a_F) is a function that assigns another function to any token *t* of that type; and that second function, in its turn, assigns the property of success to *t* exactly if a certain proposition (*that Plato snores*) *t* is true.

Unquestionably, (a_F) and (a) have exactly the same semantics. They both encode the proposition *that Plato snores.* They both ascribe exactly the same force to that proposition. Neither does anything besides this.

We may conclude that everything we just said about (a_F) applies exactly to (a). Let *t** be an arbitrary token of (a). *t** has for its semantic content a function that assigns the property of success to that same token exactly if a certain proposition is true. The proposition is: *that Plato snores.* *t** is, inter alia, a (phonetically distorted) token of "that Plato snores". A token of the latter denotes: *that Plato snores.*

*t** also comprises a (phonetically distorted) token of a force-operator. The force-operator assigns success to *t** exactly if *that Plato snores* is true. Thus, the semantic content of *t** is a function F that assigns the property of success to that token exactly if *that Plato snores* is true.

The semantic content of the sentence-*type* "Plato snores" is a function that assigns F to *t.*

Let us generalize these points.

Let *t* be *any* indicative sentence token. *t* will encode some proposition *p*. The semantic content of *t* is a function F that assigns success to *t* exactly if *p* is true.

Let *T* be the type of which *t* is a token. The semantic content of *T* is a function that assigns F to *t*.

The semantic content of an indicative sentence-type is a function that assigns such functions to tokens of that type.

The assertoric-force operator denotes a function that assigns the property of success to a token exactly if a certain proposition is true. A more exact definition of that operator must be given contextually. Given an expression token *t* that bears a proposition *p*, *t* comprises an assertoric force-operator exactly if some constituent of *t* denotes a function that assigns the property of success to *t* exactly if *p* is true.

We will develop corresponding analyses for questions and imperatives. This will involve defining the corresponding force-operators.

But before we do so, I would like to deal with a possible worry. It might seem that this account suffers from vicious circularity.

Let us consider a particular token of the expression type "Plato snores". Consider the one right below this line:

(*) "Plato snores".

That token is just an inscription – just a pattern of ink, not a pairing of ink with meaning.

Uncontroversially, there exists a function that assigns the property of success to that token just in case a certain proposition – *that Plato snores* -- is true. Obviously there is nothing wrong with that function; there is nothing circular about it. Let F* be that function.

To avoid vicious circularity, we must be careful how we state the theory. We don't want to say:

(A) The semantic content of an indicative sentence-token t is a function that assigns success to t if the proposition semantically encoded in t is true.

(A) is, or at least seems, circular. It defines the semantic content of *t* in terms of the proposition "semantically encoded in *t*". So it defines the semantic content of *t* in terms of the semantic content of *t.* Not acceptable.

What we must say is this:

(B) Part of what is encoded in an indicative sentence-token t is some proposition P. But that is not all that is thus encoded. t also comprises a function that assigns success to that token just in case P is true.

There is no circularity in (B). *t* is just an arbitrary physical (or, at any rate, spatiotemporal) entity. For any proposition P, there is nothing wrong with saying that part of what *t* encodes is P. And for any function F, there is nothing wrong with saying that F assigns the property of success to *t.* That is not the kind of function we are used to considering; but that is obviously not important. It follows that there is nothing wrong with saying that F assigns the property of success to *t* exactly if P is true.

I should clarify one other point; this will head off another false accusation of circularity. Sometimes we say things like:

The occurrence of "can" in "canary" isn't a word.

What we are saying is that a certain physical type is not appropriately paired with a meaning, and therefore doesn't constitute a word (or an occurrence thereof). So, in that case, we are using the term "word" to denote a pairing of a physical object, or kind, with a meaning. Let us say that, in cases like this, we are using "word" in a "pregnant" manner.

We don't always use "word" in that way. Sometimes we say things like:

The word "bank" has two different meanings.

Here we are saying that a certain physical type is paired with different meanings. So we are using the term "word" to denote a physical type, not a pairing of a physical type with a meaning. Let us say that, in such cases, we are using "word" in a "non-pregnant" manner.

Exactly similar remarks apply to most semantic terms: "sentence", "sentence-token", "sentence-type", "expression".

In that analysis, I am using the term "sentence-token" in a non-pregnant manner. When I talk about a "token" of the type "Plato snores", I am talking about a spatio-temporal object of some kind – an inscription, a noise, perhaps a mental image. I am not talking about a *pairing* of such an object with a meaning. If I were using the word "token" in a pregnant manner, then indeed my analysis would be circular. My analysis would then be:

(C) The content of an indicative sentence-token is a function that assigns success to that token under certain circumstances; and that sentence-token is a pairing of a physical object with a certain semantic content.

(C) defines "semantic content" in terms of "sentence-token", and it then defines "sentence-token" in terms of "semantic content". Obviously that is viciously circular. The reason it is circular is that "sentence-token" is being used in the pregnant manner.

In my analysis, I use it in the non-pregnant manner. So my analysis is not guilty of the circularity just described.

Questions and imperatives

Questions and imperatives are to be analyzed in a similar fashion. Let us start with imperatives.

A token of an imperative denotes a function that assigns the property of success to that same token exactly if the addressee makes a certain proposition be true.

If I say: "march!", my words are really elliptical for: "you: march!". So if O is the addressee, then the proposition encoded in my utterance is: *that O marches.* The "you" is semantically present, but phonetically unrealized.

To make things as clear as possible, I will always include the name of the person addressed. So I will not talk about utterances of "march!", but about utterances of "Plato, march!", "Socrates, march!", and so on.

Suppose I am talking to Plato, and I say: "Plato, march!". Let *t* be that token. Everyone who speaks English knows that, at the level of semantics, *t* is a success if it is obeyed and is a failure otherwise. Everyone who speaks English knows *which* proposition the addressee must make true if he is to obey that imperative. That proposition, of course, is: *that Plato marches.*

Again, we must distinguish semantic success and failure from other kinds. Many kinds of success can be achieved by uttering an unobeyed imperative: one can amuse, intimidate, manipulate, provoke, help, inform. And many kinds of failure can be incurred through issuing obeyed commands. But *semantically* an utterance of an imperative is a success exactly if it is obeyed. At some level, everybody who speaks a language knows that.

So, at the level of semantics, *t* is a success exactly if the addressee obeys it – exactly if, in response to my utterance, Plato makes true the proposition *that Plato marches.*

Thus, the semantic content of *t* assigns success to *t* exactly if, in response to *t*, the addressee makes true the proposition: *that Plato marches.*

So that semantic content is a function that assigns the property of success to that token exactly if, in response to that utterance, the addressee makes true that proposition.

One point should be clarified. Suppose I use an imperative to command somebody to do such and such. But suppose that person does such and such *not* because I commanded it, but for some totally unrelated reason. It seems to me that, in such a case, the imperative was a failure. This point is not central to my argument. But it seems true; and, I believe, it must be accommodated. For an imperative to be a success, it is necessary that the addressee not only make true the proposition in question, but that he do so *in response* to the imperative.

Let us now state our analysis of imperatives in a more precise way. Unquestionably, an utterance of "Plato, march!" encodes the proposition: *that Plato marches.* This makes it reasonable to suppose that such an utterance contains a constituent that refers to that proposition. So, semantically if not phonetically, "Plato, march!" comprises an expression that refers to that proposition – presumably, "that Plato marches".

It is also clear that, in such an utterance, imperative force is being assigned to that expression. So "Plato, march!" is really a phonetically-condensed or distorted form of:

(PM) "C(that Plato marches)",

where C is an operator that assigns imperative force to the expression falling in its scope. (The latter expression must denote a

proposition.)

What exactly does C do? What is it for an expression to assign imperative force to an expression?

C is *not* merely a way of expressing the speaker's attitude towards a proposition. For the sentence

(**) "I want it to be the case that you, Plato, make true the proposition: *that Plato marches.*"

has a totally different semantics from (PM), even though it expresses specifically the propositional attitude just described.

For now familiar reasons, C does not replace one expression with another. It is tempting to say:

An imperative force-operator replaces the proposition falling in its scope with some other proposition. The latter denotes the state of affairs that must obtain if the command is to be obeyed. So C is a function that takes (say) "that Plato marches" and replaces it with an expression denoting some second proposition. This second proposition is the one whose truth constitutes obedience of the command. This second proposition is: that it is the case that Plato makes true the proposition that Plato marches.

This proposal fails. Suppose that C is defined in the way just proposed. In that case, PM is synonymous with:

(**) "That Plato makes true the proposition *that Plato marches* (in response to my command)"

But (**) is not an imperative; it has no force, and is no synonym of PM.

Here is the right analysis of C. Let *t* be an arbitrary token of PM. *t* comprises a token of "that Plato marches". Of course, that expression denotes the proposition *that Plato marches.* C assigns the property of success to *t* exactly if, in response to *t*, the addressee makes true that proposition. Thus, the semantic content of *t* is a function F that assigns the property of success to *t* exactly if, in response to *t*, Plato makes true the proposition: *that Plato marches*

The semantic content of the *type* "C(that Plato marches)" is a function that assigns F to *t.*

PM is just a phonetically perspicuous version of "Plato, march". So everything we said about the former applies to the latter. An utterance of "Plato, march" has for its semantic content a function that assigns success to that same utterance exactly if the proposition *that Plato marches* is made true by the addressee in response to that same utterance.

The semantic content of the *type* "Plato, march!" is a function that assigns functions, like the one just described, to each token of that type.

Let us generalize this analysis. Let *t* be any token of any imperative sentence-type. Part of what *t* semantically encodes is some proposition P. *t* semantically encodes, and thus refers to, a function that assigns success to *t* exactly if, in response to *t*, the addressee makes it be the case that P.

Questions

Questions are dealt with in basically the same way, but there are a few wrinkles.

First of all, there are two kinds of questions: those that are to be answered with a "yes" or a "no", and those that cannot be so answered. Let us refer to the latter as "ordinary" questions.

Yes-no questions always encode a complete proposition. A token of "does Plato snore?" encodes the proposition: *that Plato snores.*

Ordinary questions encode propositional functions: proposition-like entities that contain a free variable. "Where did Smith go?" encodes the propositional function: *That Smith went to*__ There is a blank or free-variable. No complete proposition is encoded.

Let us start with yes-no questions. At the level of semantics, a question is a success only if it is answered correctly. If I ask a question, and receive no answer, then my query was a failure. This is a matter of semantics, of literal meaning; it is not a matter of pragmatics or anything else. Semantically, questions are *supposed* to be answered. Semantically, a question is a success exactly if it is answered.

But a question is not just supposed to receive *an* answer: it is supposed to receive a *correct* one. Let Smith be somebody with a peculiar cognitive disorder. He knows that "does Plato snore?" is supposed to be answered; he also knows that an answer consists in an affirmation, or a

denial, of the proposition *that Plato snores.* But Smith thinks that it is of no consequence whether the answer is correct or not; he thinks that the purpose of a question is simply to receive *an* answer, but not necessarily *the* answer.

It seems to me that Smith really doesn't understand the concept of a question. Smith really doesn't know quite how "does Plato snore?" is supposed to operate.

Further, this deficit on Smith's part is obviously semantic. It isn't that Smith has a perfect grasp of literal meaning, but falters when it comes to dealing with the nuances of implicature.

Suppose a student is irritatingly drumming his fingers on his desk, and you say, in a clearly annoyed way, "is it really necessary to do that ?" If the student is completely insensitive to pragmatic phenomena, but grasps literal meaning, he might say "technically, it isn't necessary; but I do feel like doing it, so I will carry on."

But if the student thinks that "does Plato snore?" is no more supposed to be given a correct answer than a false one, then we are no longer in the realm of implicature; we are dealing with a simple ignorance of semantics.

So a question q is a success only if it is given a correct answer. This is true both of yes-no and ordinary questions.

Of course, an unanswered question may be a great success in many respects, and an answered question may be a great failure in many respects. But *semantically* a question is a success only if it is answered.

Suppose I utter "does Plato snore?". Let *t* be that token. Everybody who speaks English – who knows what that question means – knows that, in terms of literal meaning, that question is supposed to call forth a *true* statement as to whether Plato snores. So every English-speaker knows that *t* is a success only if three conditions are satisfied. First, in response to *t*, the person addressed by that token correctly affirms or denies that a certain proposition is true. Second, that proposition must be either *that Plato snores* or *that Plato does not snore.* Third, the proposition affirmed must be true.

A couple of niceties must be dealt with.

Suppose the addressee of *t* happens to produce a sentence that gives a correct answer to that question, but he does so for some reason having nothing to do with *t*; he does so by sheer coincidence. In that case, *t* is semantically a failure.

Suppose I say "is snow white?" to you. You say "snow is white", but

you are not saying this *in response* to my question; you are saying it for some totally unrelated reason – perhaps somebody previously agreed to pay you a million dollars if you said "snow is white" at that exact moment, regardless of what anyone else was saying. My question did not prompt you to say that. So my question was, at the level of semantics, a failure. It did receive an answer, but not for the right reason.

Questions are not only supposed to be answered: they are supposed to *prompt* people to answer them.

Here, then, is our analysis.

Let *t* be an arbitrary token of "does Plato snore?" At the level of semantics, *t* is a success exactly if, in response to *t*, the addressee correctly affirms or denies the proposition: *that Plato snores.* So the semantics of *t* assign the property of being a success to that token exactly if, in response to *t*, the addressee correctly affirms or denies that proposition.

Thus, the semantic content of *t* is a function F that assigns success to *t* exactly if, in response to *t*, the addressee correctly affirms or denies the proposition: *that Plato snores*.

The semantic content of the expression-type "does Plato snore?" is a function that assigns F to *t.*

Let us generalize this point. Let *t* be any yes-no question-token you please. *t* semantically encodes some proposition P. The semantic content of *t* is a function F that assigns success to *t* exactly if, in response to *t*, the addressee correctly affirms or denies P. Let *T* be the corresponding type. The semantic content of *T* is a function that assigns F to *t.*

Ordinary questions

Let us deal with ordinary questions. Let *t* be an arbitrary token of "where did Smith go?" *t* encodes not a proposition, but a propositional function, namely: *Smith went to place x.* Let P be that function. At the level of semantics, *t* is a success exactly if, in response to *t*, the addressee produces a sentence affirming a true proposition P* that is just like P, except that P* contains a place, where P contains a variable.

Let us say that P* is a "completion" of P.

So *t's* semantic content, its referent, is a function that assigns success to that same token exactly if, in response to *t*, the addressee affirms a true proposition of the kind just described.

Let us generalize this point. Let *t* be *any* token of *any* ordinary question-type *Q*. *t* encodes some propositional function P. Semantically, *t* is a success exactly if, in response to *t*, the address affirms or denies a true proposition that is a completion of P. So *t*'s semantic content is a function F that assigns the property of success to *t* exactly if, in response to *t*, the addressee affirms or denies a true completion of P. The semantic content of Q is a function that assigns F to *t.* It is obvious how to generalize this analysis to any ordinary question.

Giving a unified semantics for questions

The answer we've given to the question "what is the semantic content of a yes-no question?' isn't quite the same as the answer we've given to the question "what is the semantic content of an ordinary question?" But "question" is not a disjunctive category. So presumably, we should be able to give an answer that applies to both yes-no and ordinary questions.

This can be done. A proposition can be considered a limiting case of a propositional function: a propositional function with zero free-variables. This is, in fact, how most logicians think of propositions. So the proposition *that Plato snores* can be considered as a kind of propositional function and, by the same token, can be considered a completion *of* that same propositional function.

Let *t* be any token you please of any question-type you please. *t* semantically encodes some propositional function P. Semantically, *t* is a success exactly if, in response to *t* the addressee affirms a true completion of P. So the semantic content of *t*, its referent, is a function that assigns success to *t* exactly if, in response to *t*, the addressee affirms or denies a true completion of P.

The general definition of "sentence"

The category of "sentence-token" is a unified one. Therefore, that category must be picked out by some one, non-disjunctive definition. What is that definition? My answer is:

(TOK) s is a sentence-token exactly if, for some proposition P, s's

semantic content is a function F that assigns success to s exactly if P is true.

Let us now define "sentence-type". First of all, a sentence-type is a platonic entity. A platonic entity is something of which there are instances or tokens. Given this point, and given how we just defined "sentence-token", we must define "sentence-type" thus:

(TYP) S is a sentence-type exactly if, for any token t of S, there is some proposition P such that the semantic content of S is a function F that assigns a function f to t such that f assigns success to t exactly if P is true.

Obviously these definitions, especially (TYP), require elucidation.

Any sentence-token *s* – whether indicative, interrogative, or imperative – is such that, for some proposition P, s is a success exactly if P holds.

Consider a token of "Plato snores". Semantically, that token is a success exactly if: *that Plato snores* is true.

Consider a token of "Plato, snore!". Semantically, that token is a success exactly if Plato is the addressee and, in response to that token, he makes it be the case that: *that Plato snores* is true. So a token *t* of "Plato, snore!" is a success exactly if the following proposition is true: *Plato is the addressee of* t *and, in response to* t, *he makes it the case that:* that Plato snores *is true.*

Consider a token of "does Plato snore?". Semantically, that token is a success exactly if, in response to that token, the addressee affirms a true proposition concerning whether *that Plato snores* is true. So a token *t* of "does Plato snore?" is a success exactly if the following proposition is true: *in response to* t, *the addressee affirms a true proposition concerning whether* that Plato snores *is true*.

It is obvious how to apply this analysis to an ordinary question, like "what did Plato eat?"

Given *any* sentence-token *s*, there is some proposition P such that *s* is a success exactly if P holds. So the semantic content of a sentence-token is a function that assigns it success exactly if some proposition holds. The

semantic content of a sentence-type is a function that pairs its tokens off with functions like the one just described.

Of course, a thing does not *just* have semantic content; it has semantic content relative to the rules of this or that language. Thus, a thing *e* qualifies as a sentence-token only if, for some language L, the semantic rules of L assign a function F to e such that, for some proposition P, F assigns the property of success to *e* exactly if P is true. A thing *E* qualifies as a sentence-type only if it pairs off its tokens with functions like the one just described.

For reasons we have just seen, if a thing *e* is to qualify as a sentence-token, it is *necessary* that, as a matter of semantics, there be some proposition P such that *e*'s success turn on whether P is true. It seems to me that it is also sufficient. Let me say why.

How are sentences different from sub-sentential expressions? Wittgenstein and Dummett answer by saying: A sentence is a "complete move in the language game", whereas a sub-sentential expression is not.[86] There is probably truth in this, but it is much too vague. What is a "language-game"? What is a "move" in that game? Many think of a language as a set of functions: How is a set of functions a "game"?

A more perspicuous answer is not hard to produce. A sentence is something that, at the level of semantics, can be *successful.* A sub-sentential expression is something that, at the level of semantics, cannot be successful. If you say simply "Plato" or "because", what you are saying cannot, at the level of semantics, be successful in any way. (Of course, I am considering tokens of "Plato" and "because" that are not elliptical for complete sentences. Obviously a token of "Plato" is successful if it is elliptical for a true sentence.) Given the semantics of English, there are no circumstances in which an isolated occurrence of "Plato" or "because" is a success. Semantically, any such occurrence is *always* a failure.[87]

But with *sentences* things are very different. Let *S* be any sentence-token of English. There is some circumstance C such that, given the semantic rules of English, S is a success exactly if C obtains.

As a matter of semantics, there is some circumstance C such that "Plato snored" is a success exactly if C obtained. Here C is the circumstance that Plato snored.

Of course, logically speaking, there is *no* circumstance C such that "triangles have four sides" is a success if C obtains. But *semantically* speaking, there is such a circumstance. The *semantics* of English leave it

open whether "triangles have four sides" can be a success or not: what prohibits success is a fact of mathematics, not of semantics. The rules of English semantics *do* assign a success-condition to that sentence: so, from the viewpoint of semantics, *there is* some circumstance C such that "triangles have four sides" is a success exactly if C obtains. That is why we can say that any such token is necessarily false: we know of the necessary non-existence of the condition under which such a token is assigned success by the semantic rules of English. So from the standpoint of English semantics, *there is* a condition under which tokens of "triangles have four sides" are a success.

The same is true of any question or imperative. A token *t* of "shut the door", addressed to O, is a success exactly if, in response to that token, O shuts the door. So *t* is a success exactly if the following proposition is true: *that, in response to* t, *O shuts the door.*

Of course, some commands cannot be obeyed. But, from the viewpoint of English semantic rules, given any imperative *t*, there is some condition C such that *t* is a success exactly if *t* is obeyed. Any token *t* of "destroy an indestructible object" cannot be obeyed. But, from the viewpoint of English semantics (though not of logic), there is some circumstance C such that *t* is a success exactly if C obtains. That circumstance is that the addressee, in response to *t*, destroy some object that cannot be destroyed. Of course, that cannot be done. But that impossibility is not *semantic* in nature. From the viewpoint of English semantics, *there is* a circumstance in which that command is obeyed and in which, thus, it is a success.

By contrast, given any *sub*-sentential expression *s*, there is *no* circumstance under which, from the viewpoint of English semantic rules, *s* is a success. (Again, I am obviously assuming that *s* is not elliptical for a complete sentence.) Of course, saying "is" in isolation might bring one great success. (Maybe someone you are wooing likes the way it sounds.) But that has nothing to do with semantics.

So it is not so wrong to define "sentence-token" thus: *S* is a sentence-token of a language L exactly if there is some circumstance C such that a semantic rule of L assigns the property of success to *S* exactly if C holds. For C to hold is for some proposition to be true. So *S* is a sentence-token exactly if, as a matter of semantics, its semantics turns on the truth of some proposition – exactly if its semantic content is a function that assigns it success just in case some proposition holds. *S** is a sentence-type exactly if it assigns functions like the one just described to its tokens.

Let us sum up. *S* is a sentence-token if, semantically, it has success-conditions. *S** is a sentence-type if, at the level of semantics, it assigns success-conditions to its tokens.

A word of caution

Our analyses of "sentence-type" and "sentence-token" are rich in quantifiers. It is very important to put the quantifiers in the right order. Consider:

(TYP) S is a sentence-type exactly if, for any token t of S, there is some proposition P such that the semantic content of S is a function F that assigns a function f to t such that f assigns success to t exactly if P is true.

(TYP) must be distinguished from the falsity:

(TYP_F) S is a sentence-type exactly if, for some proposition P, for any token t of S, the semantic content of S is a function F that assigns a function f to t such that f assigns success to t exactly if P is true.

Consider the sentence-type "I am tired". Thus there is no proposition P such that, for any token *t* of "I am tired", *t*'s success turns on the truth of P. When you token "I am tired" the success of that token turns on one proposition; when I token it, the success of *that* token turns on a different proposition. So there is no proposition P such that the type "I am tired" assigns a function F such that, for any token *t* of "I am tired", F assigns a function f to *t* exactly if P is true.

The quantifier "for some proposition P" must not come before the quantifier "for any token *t*".

If we put "for some proposition P" *after* "for any token *t*", then our definition is correct. Given any token *t* of "I am tired", there *is* some proposition P such that *t* is a success exactly if P is true. The semantic content of the *type* "I am tired" is a function F that pairs off its tokens with functions like the one just described. Thus the semantic content of "I am tired" is a function F such that, for any token *t* of that type, there is some proposition P such that F assigns a function f to *t* exactly if P is true.

A desideratum

Many think it absurd to suppose that sentence-tokens refer to anything. One reason for this, I feel, is that (indicative) sentence-tokens are true or false, while paradigmatic referring terms are not. "Plato" is not true or false. Expressions referring to propositions – for example, "that Plato snores" – are not true or false, even though the things they denote are true or false.

Sentence-tokens are true or false; paradigm cases of referring terms are neither. It follows, we are told, that sentence-tokens don't refer.

This argument is confused; it involves an erroneous conception of reference. It involves the erroneous idea that reference is severable from semantic contribution. But that argument points to a desideratum that any correct semantics for sentences must satisfy.

We observed that a *certain* circularity characterizes my analysis of sentence-tokens and sentence–types. Any analysis that satisfies the desideratum just mentioned will be characterized by a similar circularity.

Some say that sentences refer to truth-values. But it is very easy to produce an expression that uncontroversially refers to a truth-value but lacks force, and therefore isn't a sentence at all:

(t) "the truth-value *true*".

If you agree with Russell's theory of descriptions, then consider "ze truth-value *true*" or "dthat [the truth-value true]".

Let

(s) "snow is white"

be any token of "snow is white".

(t) doesn't have force; it isn't true or false, even though what it *refers* to is true or false.

Of course, those who say that (s) refers to a truth-value will say:

What is true or false is the sense of (s), not its referent.

This move is a failure. Types are what have sense, not tokens; tokens are what refer, not types. Tokens refer to the things they are assigned by the corresponding types; and tokens have no semantic content *other* than what they are thus assigned. So if we say that (s) refers to the truth-value *true*, then we are saying that it has no *other* semantic content. Thus (s) has exactly the same semantic content as a token of "the truth-value *true*". The latter is neither true nor false. So it must have a different semantic content from (s). So the idea that sentence-tokens refer to truth-values doesn't do justice to the fact that sentence-tokens are true or false.

This problem applies, though less acutely, to the theory that sentence-tokens denote propositions. Given any proposition, we can produce an expression that is neither true nor false, and that has that proposition for its sole semantic content. Let

(s*) "that snow is white"

be a particular token of "that snow is white".

Obviously (s*) refers to the proposition *that snow is white.* But (s*) isn't true or false. So it doesn't have the same semantics as (s).

Granted, (s*) and (s) *almost* have the same semantics. They both semantically encode *that snow is white*; and neither semantically encodes any other proposition. So it is *largely* right to say that (s) and (s*) refer to the same thing. But it isn't quite right.

Any theory that purports to give the semantics of indicative sentence-tokens must explain why such tokens are true or false. If we say that sentence-tokens have such and such semantic content, it is absolutely crucial that no expression having such and such content can *lack* assertoric force.

The existing theories – sentence-tokens refer to truth-values, sentence-tokens refer to propositions -- miserably fail this test.

Barwise and Perry[88] say that the semantic content of a sentence-token is a "situation" or a "class of situations." But this cannot be right. Let C be any situation, or class of situations, you please. We can produce an expression that is *not* a sentence-token, and whose referent, and whose sole semantic content, is C. Let "SC" be a name for C. "SC" isn't any more true or false than "Caesar".

Most expressions referring to functions are not true or false. This fact is especially important. For it is tempting to analyze sentence-tokens as

follows. The semantic content of a sentence-token is a function that assigns *truth* to that sentence-token exactly if a certain proposition is true. So the semantic content of (s) is some function F that assigns the property of *truth* to (s) exactly if *that snow is white* is true.

The problem is: we can easily produce an expression whose sole semantic content is F, and that isn't true or false. Here is an example. Let "F" be an expression whose sole semantic content is F. "F" is not true or false; it is no more a sentence than "Caesar". So the semantic content of "snow is white" cannot be F, at least not F alone.

Nothing that lacks assertoric force can have the same semantic content as (s), or any other sentence-token. There may be overlap, but not identity, of semantic content. Hence, the following is a desideratum that any semantic theory must satisfy:

(D) The semantic content of an indicative sentence-token must be something that an expression could not have *without* being true or false.

Any function assigning truth or falsity to (s) can be the sole semantic content of an expression that has no assertoric force. Any proposition, any fact, any truth-value, any mental state, any physical object: these can be the sole semantic content of an expression that has no assertoric force.

This is what motivates our theory. At the level of semantics, of literal meaning, (s) is a success exactly if a certain proposition is true. So (s)'s semantic content, its literal meaning, must at least include a function that assigns the property of success to (s) exactly if that proposition is true.

The semantic content of an indicative sentence-token must be a function that assigns that very token the property of success exactly if a certain proposition is true.

This theory, and no other, satisfies (D).

Let *t* be *any* expression-token you please. Suppose that *t*'s semantic content is a function phi that assigns success to *t* exactly if some proposition P is true. In that case, there is no way that *t* could fail to be an indicative sentence.

By assigning phi to *t*, you've guaranteed that, at the level of literal meaning, *t* is a success if P is true. You've guaranteed that *t* is failure if P is false. You've made it a matter of *t*'s semantics that *t* is a success just in case P is true. Thus, *t*'s semantic success turns on the truth of P. *t* asserts that P.

If you don't believe this, do an experiment. Consider the proposition *Kennedy's favorite author was Tolstoy.* (I chose this proposition because I do not know whether it is true or false.) Let "K" be some particular expression-token. So "K" is some particular bit of sound or ink or mental imagery.

Let us stipulate that "K" has a certain F* function for its semantics. F* is defined thus: F* assigns the property of success to "K" exactly if the proposition *that Kennedy's favorite author was Tolstoy* is true.

So *semantically* one thing is necessary and sufficient for "K" to be a success: the truth of the proposition *that Kennedy's favorite author was Tolstoy.*

How do we go about evaluating whether, at the level of semantics, "K" is a success? We find out if Kennedy's favorite author was Tolstoy. That is the only thing we do.

So "K" doesn't just encode the proposition *that Kennedy's favorite author* was Tolstoy. It *does* do that. But it does more. Semantically, "K" is a success exactly if that proposition is *true.*

Anyone who speaks English, and knows the semantics of "K", knows that, in terms of literal meaning, "K" and a token of

(k) "Kennedy's favorite author was Tolstoy"

are successful exactly if the proposition *that Kennedy's favorite author was Tolstoy* is true.

So our analysis correctly predicts that any token of (i) will be synonymous with "K". Our analysis makes it *impossible* for a sentence-token to have for its semantic content anything that could be the semantic content of an expression that was not true or false.

Every other analysis fails to account for this, even the theories – *especially* the theories – that identify the semantic content of sentence-tokens with the properties of being true or false. For plainly an expression can have the property of being true for its sole semantic content and not be true (or false). Consider, for example: "the property of being true".

Chapter 13 Frege on indirect discourse and compositionality

Our analysis enables us to clean up a mess that has been sullying the logico-philosophical landscape for over a century. In the process of cleaning up this mess, we will clarify important features of our own positive view.

In some contexts, replacing a sentence-token S with another sentence-token of the same truth-value preserves truth-value. Consider the sentence-token:

(1) "snow is white and grass is green"

Replacing "grass is green" with another sentence-token of the same truth-value results in a sentence-token that has the same truth-value as (1):

(2) "snow is white and Frege defined numbers as sets of sets".

A limiting case of such a context is a single sentence-token. If we replace

(3) "grass is green"

with a sentence-token of the same truth-value, what results has the same truth-value as the original.

Obviously not all contexts are like this. Consider:

(4) "Little Timmy believes that grass is green".

If we replace "grass is green" with a sentence-token of the same truth-value, what results doesn’t always have the same truth-value as the original:

(5) "Little Timmy believes that Frege defined numbers as sets of sets".

Frege concluded from this that "grass is green" has one meaning when it occurs by itself, or in "grass is green and snow is white", and a completely different meaning in "Little Timmy believes that grass is green."

In (3), "grass is green" semantically contributes a truth-value; in (4), it semantically contributes a proposition.

In general, for Frege, a sentence means one thing in truth-functional contexts, and a different thing in non-truth-functional contexts.

Frege had another reason for holding this. The meaning of a sentence is a function of the meanings of its parts. This is known as the principle of compositionality. Frege firmly believed in this principle; and, unquestionably, it is correct.

Frege held that, in (3), the semantic contribution of "grass is green" is a truth-value. He recognized that, in (4), its semantic contribution is *not* a truth-value. If Frege said that "grass is green" contributed a truth-value in (4), then he would be forced to say that the meaning of (4) was to no degree a function of what "grass is green" meant in that sentence. In other words, Frege would be forced to deny compositionality.

But Frege rightly wanted to hold onto compositionality; he wanted to say that the meaning of (4) was a function of the meanings of its parts. So he was forced to say "grass is green" means one thing in (3) and a different thing in (4).

Intuitively, Frege's analysis seems very wrong. Surely "grass is green" has the very same meaning in all of:

(3) "Grass is green"
(4) "Little Timmy believes that grass is green."
(6) "It is possible that grass is green".

In all these cases, "grass is green" contributes the proposition *that grass is green.*

Frege's analysis makes "grass is green" be *infinitely* ambiguous. It means one thing (3), a different thing in (4), yet a different thing in

(7) "Smith believes that Little Timmy believes that grass is green."

Every addition of an epistemic (or otherwise non-truth-functional) operator generates a new meaning.

But surely "grass is green" is not ambiguous at all, let alone infinitely ambiguous.

There is an objection, due to Nathan Salmon, that we must consider:

You say that "grass is green" has the same meaning in both

(3) "grass is green"

and also

(4). "Little Timmy believes that grass is green."

This is perhaps consistent with its referring to a truth-value in (3) but not in (4). Roughly, it may function as a Millian name of a proposition in (3), and as a description of a truth-value in (4). If I am right about this, then we can hold onto the idea that "grass is green" has the same meaning in both (3) and (4), while also holding that it refers to a truth-value in extensional contexts.

I'm not entirely sure I understand this objection. But what I think it means is this.

Consider some definite description, say "the teacher of Aristotle". This description can function either referentially or attributively. If you say "Little Timmy believes that the teacher of Aristotle was wise", you may be saying that Little Timmy believes something about *Plato* (referential reading) or you may be saying that Little Timmy believes that *whoever it was* that taught Plato was wise (attributive reading).

There are very good reasons for thinking that "the teacher of Aristotle" is not semantically ambiguous, and that its amenability to both referential and attributive readings is a matter of pragmatics, not semantics.[89]

The objector seems to be saying that "grass is green" is, in this respect, like a definite description. In (3) it refers to a truth-value (it refers to it by encoding descriptive information that singles it out). So in that context it is functioning "referentially". But in (4), "grass is green" refers to a proposition. So it is functioning "attributively".

But even though "grass is green" may *function* in both ways, it doesn't follow that it is semantically ambiguous. The ambiguity might be completely pragmatic, like that of "the teacher of Aristotle".

So maybe, at the level of semantics, "grass is green" always refers to a truth-value. That it sometimes *seems* not to do so is some kind of pragmatic

epiphenomenon.

Considerable support for the objection lies in the fact that, in all likelihood, "that grass is green" *is* a definite description ("the proposition that grass is green").

Evaluating Salmon's point

First of all, if "(that) grass is green" describes anything, it probably isn't a truth-value. "That grass is green" is, as we noted, nearly synonymous with "the [ze] proposition that grass is green". And the latter describes, and refers to, a proposition, not a truth-value.[90]

But let us waive this; for the sake of argument, let us suppose that "grass is green" does describe a truth-value.

Obviously "grass is green" has different truth-values in different worlds. (Even if it doesn't, just replace "grass is green" with a sentence that *does* thus vary in truth-value.) Let w be this world, and let w* be some world where English is spoken but grass is purple.

If the objector is right, then in w, "grass is green" describes the truth-value *true.* In w*, it describes the truth-value *false.*

Since English is spoken in w*, "grass is green" has the same semantics there that it has here in w.

Obviously the referent of an expression is at least *part* of its semantic content.

Therefore, in w*, "grass is green" doesn't have (quite) the same semantic content that it has here in w.

But obviously "grass is green" has exactly the same semantic content in both cases. A dilemma.

We've dealt with this sort of situation before. The way to deal with it is to say:

"Grass is green" has a two-dimensional semantic structure. In a world where grass is green, tokens of it refer to (by describing) the truth-value true. In a world where grass is not green, tokens of it refer to the truth-value false. So "grass is green" is a non-rigid designator; and whether a token of it picks out the true or the false is a function of the context (world) of utterance.

If this is right, then it is not "grass is green" *per se* that picks out a truth-value, but rather tokens of it. The expression "grass is green" *per se* wouldn't refer to anything; its semantic content would be a function of a certain kind.

A consequence is that the semantic content of a *token* of "grass is green" is a truth-value, while the semantic content of the expression-type is the just mentioned function.

Now we can see the problem with the objector's approach. If he is right, then one cannot know the semantic content of a token of "grass is green" unless one knows its truth-value in the context (world) of utterance (just as one cannot know the semantic content of an indexical unless one knows what it refers to in the context of utterance). In general, one cannot know the semantic content of a token of a sentence unless one knows its truth-value in the context (world) of utterance

But it is hard to believe that I don't know the semantic content of a token of a sentence-token if I don't know its truth-value. To understand a sentence-token – to understand a token of "that person is tired" – is to know what *would* have to be the case *if* it were true; it is not to know *whether* it is true. The most natural position, I think, is to say that one *can* understand sentence-tokens without knowing their truth-values; and, consequently, that such tokens do not refer to truth-values.

As we've seen, it is not easy to make the case that sentence-*types* refer to truth-values. For obviously the truth-value associated with the sentence-*type* "grass is green" varies from world to world, and possibly even from time to time, even though its semantics does not thus vary. So unless we are willing to take the heroic measure of saying that the referent of an expression is not part of its semantic content, we should distance ourselves from the idea that sentence-types refer to truth-values.

Indirect discourse: a positive account

Frege *did* make a point about sentence-reference that, I think, is quite correct and that forms the basis of a correct analysis of indirect discourse. He said that, in

(4) Little Timmy believes that grass is green,

"grass is green" refers to the proposition: *that grass is green.*

A correct account of indirect discourse merely generalizes this point of Frege's. It extends it to contexts where, if we over-value phonetics, it seems not to apply.

Let us consider our paradigms:

(i) "Plato snores".
(ii) "Fred believes that Plato snores."
(iii) "It is possible that Plato snores."
(iv) "that Plato snores"

Let (i)-(iv) be expression-tokens, not expression–types. Obviously each of (i)-(iv) semantically contributes *that Plato snores.* To some degree, though not entirely, this is consistent with their phonetics. For reasons earlier discussed, we may assume that, semantically if not phonetically, they all comprise occurrences of "that Plato snores". I propose that "that Plato snores" denotes the same thing in all of (i)-(iv): the proposition *that Plato snores.* What varies from case to case is what is being said *about* that thing. Very roughly, what is happening is this.

(i) comprises phonetically unrealized material that ascribes truth to that proposition. Consequently, (i) is true just in case "that Plato snores" denotes a true proposition.

(ii) comprises material that ascribes the property of being believed by Fred to that proposition. Consequently, (ii) is true just in case that proposition has that property.

(iii) comprises material that ascribes the property of being possibly true to that proposition. Consequently, (iii) is true just in case that proposition has that property.

(iv) doesn't comprise any material other than that expression; it doesn't ascribe any properties to that proposition or to any other proposition. Consequently, (iv) isn't true or false.

So there is no constituent of (i) that has one meaning in (i) and a different one in (ii) or (iii) or (iv). Nothing has shifted meaning. The only

shift relates to what has been *said* about the thing denoted by "that Plato snores".

"Plato snores" occurs truth-functionally in (i), and non-truth-functionally in (ii), because in (i) truth has been ascribed to the corresponding proposition, whereas in (ii) truth has *not* been ascribed to it.

Here is another way to think about it. Consider the sentence-tokens:

(i_T) "*That Plato snores* is true."
(ii_T) "*That Plato snores* is believed by Fred.
(iii_T) "*That Plato snores* is possible."

Obviously "that Plato snores" means the same thing in all of these cases. Also, it is clear that (ii_T) is a mere re-wording of (ii), and that (iii_T) is a mere re-wording of (iii). Obviously "that Plato snores" occurs truth-functionally in (i_T). And, clearly, it does not mean something different from what it means in (ii_T) and (iii_T). The fact that the one context is truth-functional, while the other two are not, doesn't have anything to do with a meaning-shift. So there is no *general* inference from "S is occurring non-truth-functionally" to "S has undergone a shift in meaning."

Obviously (i_T) is very close to (i). It seems more natural to say that (i) is a compressed form of (i_T) than it is to say that the words "Plato snores" have entirely different meanings. (This is not *quite* accurate: see below.)

Also, it isn't good methodology to posit meaning-shifts every time a context is non-truth-functional. "Kennedy died" occurs non-truth-functionally in "Kennedy died because he was shot", "Kennedy died after Lincoln died", "it is regrettable that Kennedy died", and so on. Most connectives are non-truth-functional. The only ones that are truth-functional are "not", "and", "or", and "it is true that". (I am leaving aside manufactured ones like "it is true that it is true that", and "it is not the case that it is not the case that".) There is no need to posit meaning-shifts, as long as we do not put too much stock in phonetics.

We can be more precise about this. Semantically (i) is a phonetic variant of:

(i_F) "A(that Plato snores)",

where "A" is the assertoric force-operator. By similar reasoning, (ii) and (iii) are to be rewritten thus:

(ii_F) "A(that Fred believes that Plato snores)."

(iii_F) "Q(that Plato snores). "

All of (i_F)-(iii_F) are meant to be sentence-tokens, not sentence-types. We've already discussed how to analyze such tokens. For example, (i_F) comprises an occurrence of "that Plato snores", and thus encodes the corresponding proposition. The "A" denotes a function that assigns the property of success to (i_F) exactly if *that Plato snores* is true. So, in effect, the "A" in (i_F) serves the function of ascribing truth to the proposition *that Plato snores.* (It doesn't do this as directly as might be thought.) Similar remarks apply to (ii_{iF}) and (iii_F).

Let us extend these points. Connectives like "after", "because", "it is regrettable that", "Bill is happy that" are non-truth-functional. Most connectives are non-truth-functional. We deal with this not by positing meaning-shifts, but by filling in some semantic gaps that phonetics leaves us. Consider:

(vi) "Kennedy died because Kennedy was shot".

Let (vi) be a token, not a type.

(vi) is true exactly if both constituent sentences are true, and the truth of the second is causally responsible for the truth of the first. So (vi) is equivalent to:

(vii) Because* <that Kennedy died, that Kennedy was shot>,

where "because*" denotes a function assigning truth to an ordered pair exactly if that ordered pair comprises two expressions denoting true propositions, and the truth of the first proposition is a consequence of the truth of the second.

For analogues of reasons discussed earlier, I would suggest that (vii) is a mere phonetic variant of (vi).

But (vii) does not make the semantics of (vi) as perspicuous as possible. (vii) must be rewritten thus:

(vii_F) A(that because* <that Kennedy died, that Kennedy was shot>).

Let (vii_F) be a token, not a type.

The expression falling in the scope of the "A" denotes a proposition. (vii_F) is true exactly if that proposition is true.

(vii_F) makes perspicuous the semantics of (vi).

I would suggest that sentence-tokens containing the connectives "after", "it is regrettable that", "Bill is happy that" are to be analyzed in an exactly similar way.[91]

A slight inaccuracy

I said above that

(i) "Plato snores"

is synonymous with:

(i_T) "*That Plato snores* is true."

My larger point is that "Plato snores" has the same meaning in (i) that it has in "Fred believes that Plato snores". And if we see (i) as being a phonetic variant of (i_T), the motivation for that claim becomes clear.

But, technically, (i_T) isn't quite synonymous with (i). According to our theory, (i) is a mere phonetic variant of:

(i_R) "A(that Plato snores)"

But (i_T) is a phonetic variant of:

(i_{TF}) "A(that *that Plato snores* is true)."

(i) and (i_T) aren't quite synonymous. But this obviously doesn't undermine our larger point.

Am I over-using the concept of deep-structure?

I should deal with a misgiving some may have about my analysis:

You talk very freely about "phonetically realized" sentence-constituents and, by implication, about "surface-structure" and "deep structure". But it is quite possible that all such talk is hogwash. Granted, Chomsky and others talk and think this way. But it is very much an open question whether those people are right. Your commitment to Chomsky's dubious paradigm weakens your analysis.

This criticism has some force. But let me make a couple of remarks in my defense.

I leave it open whether Chomsky is right. But I don't think my analysis requires that he be right. My analysis has its roots in the very *concept* of a linguistic expression; its truth isn't contingent on the fortunes of an empirical doctrine like Chomsky's.

The English word "Socrates" can be tokened by physical entities that have absolutely nothing in common. First of all, that word can be spoken or written or expressed by hand-signals. No hand-signal has *any* physical similarity to any spoken token or any written token. No written token has any similarity to any spoken token.

Further, given any two written tokens of "Socrates", they may bear minimal resemblance to each other. If a court-reporter tokens that word in his short-hand, the result will bear little, if any, resemblance to what I write if I type it or if I write it in cursive.

If a person with damaged vocal chords tokens that word, the result bears no resemblance to what results when I token it.

What makes a physical entity count as a token of a certain word is

not merely its physical properties – is not its shape, or its acoustical properties. Such properties have a role, but not the role one might think.

A corollary is that if an inscription perfectly isomorphic with the following

SOCRATES

were tokened, but it weren't tokened in the right cultural context, or in consequence of the right intentions, then it wouldn't be a token of the word "Socrates".

Suppose a coffee-machine were by sheer accident to produce the sounds "Socrates was wise". Would the machine have *spoken*? Would it even have produced any tokens of any words? No. It would have produced sounds that, had *we* produced them, would have qualified as speech. The machine hasn't spoken at all; it hasn't said anything. It made sounds, but it was not using those sounds to execute some linguistic intention. So it didn't produce word-tokens. It produced sounds that, had we produced them, would have been words. Where the right psychological background is lacking, there are no word-tokens.

But the right psychological background is not enough. First of all, one never just speaks or writes; one always speaks or writes in this or that language. Now suppose I *think* that, in English, "blrak blurg gralb" means *grass is green*; and I say "blrak blur gralb" with the intention of saying *grass is green.* Because there is no existing practice of taking "grass is green" to mean *grass is green*, I haven't said anything, at least not in English. And I haven't said anything in any other language. As we saw earlier, when discussing "Sukrat", if a noise that I produce is to qualify as a token of a word of some language L, then that noise must be a consequence of my believing that *in L* that noise has a certain meaning. My utterance of "blrak blur gralb" was a consequence of my belief that *in English* that noise means *grass is green*; it was not a consequence of any beliefs about L. So I didn't say *anything* in any language.

What makes a physical event be a tokening of this or that expression is not that it has these or those phonetic properties. It is that the event is caused by certain psychological events, in a certain cultural or historical context.

The phonetic or geometrical properties of that event help others to

identify the presence of the psychological and cultural factors that allow for the tokening of expressions. That seems to be their primary function. Up to a point, those phonetic properties may even be more than just indicators of such factors; up to a point, phonetics may be constitutively involved. But the presence of those phonetic properties is not, at least not by itself, what *makes* that event be an actual expression-tokening, as opposed to something that just sounds or looks like one. Phonetics has a role; but it is not as central as might be thought.

Any expression can be physically realized in different ways. An expression-type like "that grass is green" is no exception. For the type "that grass is green" to be tokened is, at least in part, for something with a certain *semantics* to be tokened. What gives a physical event the right semantics is, at least in part, that it has the right psychological and cultural background. If that background is present, there is enormous latitude as to what the actual physical properties – the phonetics, the morphology – of that token can be. My theory asks that "that grass is green" have different phonetic (and acoustic) incarnations. In light of what was just said, this is not much to ask.

Chapter 14 Opacity revisited: the inadequacy of Kaplan's analysis, the viability of other analyses

Everybody thinks it *strange* to suppose that "grass is green" means one thing on its own and a different thing in "John believes that grass is green". But most believe this supposition to be coherent.

There are a couple of reasons to doubt this.

A Kripkean reason

Frege says that certain operators cause sentences falling in their scope to undergo a shift in meaning. A consequence of this position is that repeated application of such an operator induces repeated shifts in meaning.

Frege's position is inconsistent with Kripke's discovery that, at the level of semantics, proper names do not have senses.

According to Frege,

(i) "Bertrand Arthur William Russell is Bertrand Arthur William

Russell"

contributes a truth-value. In

(ii) "Fred believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell".

"Bertrand Arthur William Russell is Bertrand Arthur William Russell " contributes a proposition: the proposition it expresses in (i). When it occurs in

(iii) "Jerry believes that Fred believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell",

"Bertrand Arthur William Russell is Bertrand Arthur William Russell " contributes a third thing: the proposition it *expresses*, not the one it refers to, in "Fred believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell".

Frege's theory requires that, in (ii), "Bertrand Arthur William Russell is Bertrand Arthur William Russell" have both sense and reference. Its reference in (ii) is a proposition: *that Bertrand Arthur William Russell is Bertrand Arthur William Russell.* Its sense in (ii) is some proposition *through* which we grasp its referent.

We know from Kripke that proper names of individuals are not semantically associated with senses. Semantically "Bertrand Arthur William Russell" does not have a sense.

Frege's theory requires that "Bertrand Arthur William Russell" have a sense. Otherwise, that theory falsely assigns *no* meaning to oblique occurrences of "Bertrand Arthur William Russell". Since "Bertrand Arthur William Russell" doesn't have the required sense, Frege's theory fails.

Further, Frege's theory requires that "Bertrand Arthur William Russell" be *infinitely* ambiguous in two quite distinct ways. First, it requires that it mean one thing in (i), another thing in (ii), a third thing in (iii), and a fourth thing in

(iv) "Fred believes that Bob believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell".

So Frege’s theory makes "Bertrand Arthur William Russell" infinitely ambiguous in that way.

But we must remember that any object can be given by infinitely many different modes of presentation or senses. There are infinitely many descriptions that single out anything. So, if Frege’s theory is right, then "Bertrand Arthur William Russell" is infinitely ambiguous in (iii), infinitely ambiguous in (iv), and so on.

But "Bertrand Arthur William Russell" isn’t ambiguous at all, let alone doubly infinitely ambiguous.

Even if, by some remote chance, two people are named "Bertrand Arthur William Russell", so that it is ambiguous, it still wouldn’t be ambiguous in the way required by Frege’s theory.

The problems with Frege’s theory are well documented. But what is not realized is that they bedevil any theory according to which non-truth-functionality involves meaning-shift.

Suppose you say that, in truth-functional contexts, sentences refer to truth-values. You must deal with the fact that they obviously *don’t* refer to such things in other contexts. The most plausible way to deal with this is by supposing, as Frege did, that certain operators induce meaning-shifts. If you go this route, you must answer the question: What is the nature of that shift? What does the sentence refer to before the shift, and what does it refer to afterwards?

The most plausible answer is the one Frege gives: *after* the shift, the sentence refers to the thing that, prior to the shift, it had for its sense. But we’ve just seen the problems with this view.

A Davidsonian reason

Donald Davidson[92] put forth a clever argument that challenges Frege’s analysis of indirect discourse. I do not know whether Davidson’s argument is decisive. But it is powerful and deserves to be mentioned. [93] [94]

English is obviously a learnable language. Any learnable language has a finitely large primitive lexicon. If the Frege-Church line is correct, then "snow is white" means one thing on its own, a second thing in "Tim said that John believes that snow is white", and so on.

You cannot understand *all* of these sentences. But that has nothing

to do with your *semantic* knowledge. It has to do with your ability to *exploit* your semantic knowledge. It has to do with matters relating to fatigue, limitations of memory, and general cognitive ability. If we considered a being that had exactly your semantic knowledge, but who had unlimited energy, logical prowess and so on, that being *would* be able to understand an occurrence of "snow is white" that was pre-fixed by 42,345 epistemic operators. If you could fully exploit your knowledge English semantics, you could assign meaning to such a sentence. Any semantic theory must be consistent with this fact.[95]

Let "snow is white$_1$" be synonymous with the sentence "snow is white" when it occurs on its own, or in an otherwise truth-functional context. Let "snow is white$_2$" be synonymous with "snow is white" when the latter occurs in the scope of a single epistemic operator. Let "snow is white$_3$" be synonymous with "snow is white" when the latter occurs in the scope of two epistemic operators. And so on.

Anyone who speaks English has the semantic resources to understand "snow is white$_{7,865}$". If Frege is right, for any i, your knowledge of English semantics comprises a knowledge of one semantic rule for "snow is white$_i$", and a different one for "snow is white$_{i+1}$". So your knowledge of English semantics comprises *infinitely* many semantic rules, corresponding to "snow is white".

By itself, this might seem acceptable. After all, English semantics *does* comprise infinitely many rules. But most of them are *derived*; and one can learn those on the basis of the primitive rules. If Frege's theory is right, then English comprises infinitely many *primitive* semantic rules.

Frege's theory, by implication, makes "snow" be *infinitely* ambiguous. There is "snow$_1$", "snow$_2$", "snow$_3$", and so on. According to that theory, "snow" is ambiguous between all these different words.

For each i, "snow$_i$" is a *primitive* entry in our lexicon. It has no constituent structure; no recursive rules give it meaning.

But, as we've seen, every English speaker *does* know, for any number n, what is meant by "snow" when it is prefixed by n epistemic operators. So Frege's analysis requires that each person who speaks a natural language know *infinitely* many primitive semantic rules. This is not acceptable. It makes English unlearnable. Besides that, it simply isn't plausible: English doesn't comprise infinitely many primitive semantic rules.

Granted, this argument is not airtight. One could make the following objection to it:

Frege does make "snow" infinitely ambiguous – there is "snow$_1$", "snow$_2$", and so on. But there is not a separate, primitive rule for each of these. There is one recursive rule, namely: "snow$_1$" means snow, and for i, "snow$_{i+1}$" refers to the sense of "snow$_i$".

Oddly, Davidson doesn't consider this.

In any case, this objection is easily met. For the sake of argument, let us suppose that, for each i, "snow$_i$" semantically encodes a sense. This supposition is false. But without it, Frege's argument fails. So let us grant it.

There is "no backwards road" from denotation to sense, Russell[96] once said. Any given thing can be given by infinitely many different senses. After all, any thing can be uniquely singled out by infinitely many different descriptions.

So, for each i, there are infinitely many possible senses that "snow$_i$ " could have. The recursive rule just described would have to choose some *one* sense. Any such choice would obviously be arbitrary; it wouldn't correspond to any fact about the actual semantics of "snow". Further, the semantic rule described would have to select a single member from each one of an infinitely large number of infinitely large classes. It is, to my knowledge, an open question in set theory whether such a rule could exist.

I conclude that Davidson's argument has some force, and that attempts to meet it redound to the discredit of Frege's theory.

For reasons discussed earlier, Frege's view is not an anachronism. If we say that sentences refer to truth-values, then we must probably adopt Frege's view. We must say that "snow" is infinitely ambiguous, and that each of its disambiguations is given meaning by a primitive semantic rule. So, if we say that sentences refer to truth-values, we are saying, by implication, that English comprises infinitely many primitive semantic rules.

Chapter 15 Our semantics

Given what we've said in this work, certain reasonable-seeming semantic views must be adjusted.

In order to give a unified view of our semantic results, we must first deal with one problem. Part of what we are trying to do is to give an analysis of noun-phrases. Among such phrases are different kinds of existential quantifiers: "something", "some man", "three birds". According to Russell, definite descriptions are existential quantifiers.

Frege has given us a very plausible analysis of such expressions. But a commonly accepted analysis of them actually makes Frege's analysis be wrong; more exactly, it makes Frege's analysis be only *partially* correct. I believe that Frege's analysis is unqualifiedly correct. So I must show why this other analysis, due to Quine[97], is wrong.

Quine on "some"

According to Frege, "some" and "for some x", and the like, are to be defined contextually:

(F) "…some phi…" means: *there is some x such that x has phi and… x…*

In some cases, the expression "a phi" is to be analyzed in the same way. Sometimes "a" means *any*: "a man is only as good as his word". But sometimes it means "some": "a man stole my wallet". We will focus on the second meaning.

As Quine[98] pointed out, some sentences of the form "…a phi…" (or "…some phi…") are ambiguous between wide-scope and narrow-scope readings. On the one hand,

(i) "John wants a sloop"

can be taken to mean that there is some *particular* sloop that John wants. On the other hand, (i) can be taken to mean that John wants mere "relief from slooplesness", as Quine put it.[99]

An exactly similar ambiguity is generated whenever an occurrence of "some" falls within the scope of an epistemic operator like "believes" or "is looking for". Let us initially discuss "wants"; we will then generalize our

analysis to these other operators.

According to Quine, the first reading of (i) is this.

(i_W) There is some x such that x is a sloop and John wants x.

The second reading, says Quine, is this:

(i_N) John wants: that there is some sloop x such that John has x.

I have only one disagreement with Quine. Notice that in (i_W), the complement of "wants" is an *object*, whereas in (i_N) the complement of "wants" is a *proposition.* If that is right, then in some cases what is wanted is a *thing* – a vase, a sloop, a car -- whereas in other cases what is wanted is that a certain *proposition* be true: one involving a vase, a sloop, a car. So on Quine's analysis, "wants" is ambiguous between "wants$_T$" and "wants$_P$". The former takes objects for its complements; the latter takes propositions. The latter denotes a propositional attitude; the former does not – it denotes an "objectual attitude".

I doubt that "wants" is ambiguous, semantically or even just pragmatically, between "wants$_T$" and "wants$_P$". I am pretty sure that the kind of thing "wants" denotes in (i_W) is identical with the thing it denotes in (i_N). I am pretty sure that "wants" *always* denotes a propositional attitude.

Also, even if we leave aside matters of semantics, I really doubt that the very concept of desire is thus disjunctive: I doubt that there is one kind of wanting that concerns states of affairs and another that concerns objects. I may "want" a particular sloop; I may "want" *some* sloop or other, not this or that specific sloop; but what varies is not the nature of wanting, but the identity of the thing I want.

Here is what I would propose. In *every* case, the object of "wants" is a proposition. When we say that so and so and "wants" x, we are really saying that so and so wants some proposition to be true concerning so and so's relation to x. If I say: "Smith wants a car", I am saying: Smith wants it to be the case that: Smith has a car.

The story is not relevantly different if there is some *specific* car that Smith wants. Suppose we say: " there is some *particular* car x such that

Smith wants x." What we are saying is: " there is some *particular* car x such that Smith wants that: Smith has x."[100]

So Quine's *narrow*-scope reading of (i) is quite correct. It is his *wide*-scope reading that is wrong. The right wide-scope reading of (i) is not (i_W). It is:

(ii) There is some particular sloop x such that John wants: that John has x.

I would like to briefly discuss the difference between *semantic* and *syntactic* ambiguity. Any adequate semantics for intensional verbs *requires* them to *be syntactically*[101] ambiguous. At the same time any such semantics must not make them *semantically* ambiguous. Quine's analysis makes them semantically, and not just syntactically, ambiguous.

What do these terms mean?

Every student of logic is familiar with the sentence:

(a) Every man loves a woman.

There are two readings:

(b) There is some woman x such that, for any man y, y loves x.

(c) Given any man x, there is some woman y, such that x loves y.

(b) says that some particular woman is loved by all men. (c) does not say this.

So (a) is ambiguous. But it is not *semantically* ambiguous. No constituent has two readings. Each constituent has one reading, but these constituents can parse out in different ways.

Another famous example of syntactic ambiguity is:

(d) "it is not likely that any given person will win the lottery".

This is ambiguous between

(e) For any x, it not likely that x will win the lottery

and

(f) It is not likely that, for any x, x will win the lottery.

(e) is true; (f) is false. (d) is ambiguous. None of its constituents has more than one reading. But those constituents can parse out differently; so (d) is syntactically ambiguous.[102]

In light of this, consider the sentence-token:

(i) "John wants a sloop".

Quine says that this is ambiguous between:

(i_W) There is some x such that x is a sloop and John wants: x.

and

(i_N) John wants: that there is some sloop x such that John has x.

(i_N) attributes a *propositional attitude* to John. (i_W) does *not* attribute any propositional attitude to John; it attributes an "objectual" attitude to John - an attitude towards an object.

There are two reasons this is not acceptable. First, if you say that John "wants" anything, you are *always* attributing a propositional attitude to

him, even if we give a de re reading to the statement in question.

Suppose John wants some *specific* sloop – say, *The Winds of War*. What John wants is that a certain *proposition* be true, namely: the proposition *that John owns The Winds of War*.

Even if you want some *specific* woman – Sally, say -- to be your bride, you still have an attitude towards a proposition; you want the proposition *that Sally is my bride* to be true. You don't just want: *Sally*.

In general, even when it is a *specific* thing that one wants, one has a propositional, not an objectual, attitude.

That by itself shows that Quine's analysis is wrong. But even if we waive the last point, Quine's analysis still fails. If "wants" sometimes denoted a propositional attitude, and sometimes did not, then "wants" would surely be *semantically* ambiguous: how could an expression that denoted two entirely different things fail to be? But "wants" is not semantically ambiguous, at least not in the way demanded by Quine's theory.

For reasons exactly similar to those just stated, Quine's analysis makes *any* intensional verb be semantically, and not just syntactically, ambiguous. It is pretty clear that his de dicto reading is right. That reading does justice to the fact that, when John wants something, he has an attitude towards a proposition. It is Quine's de re reading that is wrong.

We must replace Quine's de re reading with the one we have proposed:

(ii) There is some particular sloop x such that John wants: that John has x.

This reading does justice to the fact that, whenever John wants something, he has a propositional attitude. Also, unlike Quine's reading, our reading does *not* render "wants" semantically ambiguous. Our "wants" always denotes the same propositional attitude.

Of course, the specific proposition that John wants to be true will vary. Sometimes he will want the truth of: *John owns some boat or other.* Sometimes John will want the truth of: *John owns the Queen Mary.* But "wants" always denotes a propositional attitude; it doesn't shift between objectual and propositional meanings.

Analogues of Quine's analysis render *all* intensional verbs semantically ambiguous. Analogues of the analysis we just gave appropriately undo that ambiguation.

The relevance of this to our project

According to Frege, (F) is the semantic rule for "some phi". But it *seems* that when we apply (F) to (i), what we produce is never (i_N); it *seems* that, to generate (i_N), we cannot apply (F) in a purely mechanical way, but must make *ad hoc* adjustments. This would suggest that (F) is not quite accurate.

At the same time, (i_N) is, I have no doubt, a legitimate reading of (i). Obviously not *all* legitimate readings of (i) are de re readings. Obviously intensional verbs are often to be given de dicto readings.

My point is that none of this is any threat to the correctness of (F). (F) is correct. Applied in a purely mechanical way, without any *ac hoc* adjustments being made, (F) always yields the right result when applied to (i), or to other sentences containing existential quantifiers falling in the scope of intensional verbs. To see this, we must realize that (i) is really a compressed form of:

(i_C) "John wants: that John has a sloop.

(i_C) is indeed ambiguous. But (F) generates exactly those two disambiguations and no others. When we apply F, what results is either:

(ii_{CN}) John wants: that there is some x such that x is a sloop and John has x

or

(ii_{CW}) There is some sloop x such that John wants: that John has x.

Both readings are generated by (F). The way that Quine reads (i)

makes (F) turn out false, at least on one legitimate reading. Quine's reading has the consequence that (F) is right for the wide-scope/de re reading, but wrong for the narrow-scope/de dicto reading. But (F) is right for both readings. We must simply remember that "wants", and other intensional verbs, unfailingly have *propositional* complements; and we must remember that phonetics sometimes obscures this fact.

So before we apply (F) to (i), we must undo the distortions of phonetics. We must rewrite (i) thus:

(i_p) "John wants: that John has a sloop".

When we apply (F) to (i_p), we generate two readings: and *both* are correct:

(i_{PN}) "John wants: that there is a sloop x such that John has x."
(i_{PW}) "There is a sloop x such that John wants: that John has x.

Thus (F) gives the right reading across the board.

As we saw earlier, if (F) is right, then "some" (and "a") can be seen as denoting a function that assigns truth to a class exactly if that class is instantiated. We can thus accept that definition, and know that no modification of it is demanded by contexts like (i).

We saw how Quine's analysis partially falsified Frege's treatment of "some phi"; it makes Frege's analysis produce the right reading in only *some* cases. If we extend Quine's analysis to other quantifiers, it yields exactly parallel falsifications of the time-tested analyses of them as well. Consider the sentence

(b) "John wants every boat".

If we apply Quine's analysis to (b), we end up with:

(b_N) John wants: that for any boat x, John has x.

and

(b_W) For any boat x, John wants: x.

If this natural extension of Quine's analysis is correct, then "wants" denotes a propositional attitude in the de dicto reading of (b); and in the de re reading, it denotes something that is *not* a propositional attitude – it denotes an "objectual" attitude. But this is absurd. Obviously "wants" always denotes a propositional attitude. So Quine's analysis fails.

More importantly, that analysis falsifies the standard analysis of "every phi". The standard analysis is:

(e) "…every phi…" means that: for any x such that x has phi…x….

(e) is an algorithm: a mechanical way of assigning meaning to sentences of the form "…every phi…" If Quine's analysis of (b) is correct, then only *one* reading of (b) is generated by (e). That reading is (b_W). The problem is that (b_W) is not even *a* legitimate reading of (b); for in (b_W) the verb "wants" doesn't denote a propositional attitude. So Quine's analysis, extended to "every", makes (e) be quite false.

But (e) is a time-tested and fruitful analysis. If Quine's analysis were right, we would have to make drastic changes to our current semantic theories.

Fortunately, we don't have to do this. What is false is not (e), but Quine's analysis. The verb "wants" always has a propositional complement. The same is true of any intensional verb. Before we apply (e) to a sentence containing such a verb, we must take care to make its semantics coincide with its phonetics:

(b) John wants: that John has every boat.

If we apply (e) in a completely mechanical way to (b), we generate two readings:

(b_{WK}) For any boat x, John wants: that John has x.
(b_{NK}) John wants: that for any boat x, John has x.

On this analysis, "wants" denotes a propositional attitude on the de re reading: this is as it should be, since "wants" *always* denotes such an attitude. Further, "wants" denotes the very same propositional attitude on the de dicto reading. Again, this is as it should be: for "wants" is not semantically ambiguous, at least not in (b).

Of course, (b_{WK}) and (b_{NK}) ascribe different propositional attitudes to John. But this is not because "wants" has different meanings in them. It is because different propositional functions are the verbal objects of "wants". Quine's analysis wrongly has "wants" denoting altogether different kinds of thing.

Most important of all, on our analysis, (e) is unqualifiedly right: it *always* gives the right meaning.

If (e) is right, then "every phi" denotes a function that assigns truth to a class C exactly if C is universally instantiated.

Since (e) is right, we can accept that analysis.

What we've said about (e) and (F) is true of many other quantifiers: "three men", "two girls", "most rabbits". There are time-tested and fruitful analyses of these quantifiers. When we extend Quine's analysis to occurrences of them that fall within the scope of intensional verbs, these analyses turn out to be falsified, for reasons exactly parallel to those given above. We undo this damage by rejecting these extensions of Quine's analysis, and thus by rejecting Quine's original analysis; this involves realizing that intensional verbs always take propositional complements, even when given de re readings.

Our semantics

Now we can give a unified semantics for noun-phrases.

We say that expression-types don't refer to anything. So the type "for some x" doesn't refer to anything, not even a function. But tokens of it do.

So if *t* is a token of "for some x", then the semantic content of *t* is a function F that assigns truth to a class C exactly if C is instantiated.

Let *T* be the type corresponding to *t.* The semantic content of *t* is a function that assigns F to *t.*

Exactly similar remarks apply to "some" and (mutatis mutandis) to

"some man", "something", and so on.

It is pretty clear how this analysis generalizes to other quantifiers. If *t* is a token of "nothing" ("everything", "most things", "many things"), then *t* denotes a function F that assigns truth to a class C exactly if C is uninstantiated (universally instantiated, usually instantiated, instantiated to a non-negligible degree). The semantic content of *T*, the corresponding type, is a function that assigns F to *t.*

Obviously "some man" is in the same *grammatical* category as "Plato" and "that man", "the author of Waverly", and "ze author of Waverly".

Intuitively, we want to say that there is some significant and non-disjunctive property had in common by all and only those things we call "noun-phrases". Semantics has to respect grammar, at least up to a point. Grammatical categories may be useless when it comes to logic or metaphysics. But they cannot be useless when it comes to semantics.[103] How could they be? When you say that x, y, and z are "adjectives" or "determiners" or "nouns", you are saying something specifically about the kinds of contributions that x, y, and z can make to the meanings of sentences in which they occur; you are saying something specifically about the *semantics* of such expressions. So it is a truism that semantics, at least the semantics of natural language, must respect grammar.

We don't want to say that (tokens of) *some* noun-phrases denote second-level functions, while others denote individuals. That, it seems to me, would make "Plato" fall into a very different grammatical category from "some man" and "that man".

This last point, I grant, is not exactly air-tight. But what is clear is that *if* "Plato", "some man", "all birds" all denote the same kind of thing – if they all denote individuals or they *all* denote functions – then it will be a lot easier to do justice to the fact that they are all in the same grammatical category than it would be if we said that some of them denote an individual, while others denote, not just functions, but second-order functions (functions that take functions as arguments).

The problem is that "some man" does not denote *any* individual. No quantifier does. There is no doubt about this. Inevitably, some say that "some man" denotes an arbitrary individual[104] or an individual in somebody's "belief-world".[105] But Russell's (1920) arguments against this view seem decisive.

At the same time, we know from Kripke (1972) that "Plato" is not

the kind of quantifier that Russell thought. It is hard to believe that "Plato" is anything other than a term that contributes Plato. Thus, "Plato" denotes Plato, and not a function: and once again we have the problem of giving a unified semantics for "Plato", "some man", and so on.

I think there may be a solution.

A token of "Plato snores" is true exactly if the class of snorers comprises Plato. So "Plato" can be seen as denoting a function that assigns truth to a class exactly if Plato is a member of that class.

Thus a token of "Plato" denotes a function that assigns truth to a class C exactly if Plato is a member of C.

Tokens of "Plato" are thus in the same basic category as tokens of "some man", "no man", and so on. They all denote functions that assign truth-values to classes.

Tokens of "Plato" are not exactly *quantifiers.* A quantifier is an expression that denotes a function assigning a truth-value to a class on the basis of *how many* things of a certain type are in that class. That is not what tokens of "Plato" do.

But that doesn't threaten the semantic unity we have established. All of "Plato", "some man", and so on, denote the same basic kind of function.

Our analysis is consistent with Kripke's insights. On our analysis, "Plato snores" is true exactly if *Plato* snores – it isn't necessary that a teacher of Socrates or an author of the *Republic* be a snorer. So tokens of "Plato" are directly referential. But they are directly referential by way of encoding functions of the kind just described.

The semantic content of the expression-type "Plato" is a function that assigns the functions just described to tokens of "Plato".

We know from Kaplan (1989) that tokens of indexicals are directly referential. For example, a token of "you", addressed to so and so, semantically contributes so and so himself, as opposed to a concept or description that applies to so and so. If I address Smith, and I say "you are looking well today", the proposition encoded in my words is: *Smith is looking well today*. The token of "you" is *directly referential*: it contributes Smith, not a description that applies to Smith. In general, tokens of "you" are directly referential.

In my view, this is correct. But it must be adjusted. We want such tokens to be in the same general category as "Plato" and "some man". So the right rule is this. A token of *t* "you", addressed to x, denotes a function F that

assigns truth to a class exactly if x is a member of that class. If there is no addressee, the utterance is abortive. The semantic content of *T*, the corresponding type, is a function that assigns F to *t*.

If Russell is right about definite descriptions, then for any token *t* of "the phi", *t* denotes a function F such that, for any class C, F assigns truth to C exactly if for some x, x is a phi that falls in C, and no phi falls outside of C. The type *T* is the function that assigns F to *t*. (I am ignoring niceties relating to contextual salience.)

If the referentialist is right, then "the phi" is defined thus. If it is the case that, for some x, x is a unique phi, then a token *t* of "the phi" denotes a function F that assigns truth to a class exactly if C comprises x. If there is no such x, then *t* denotes nothing – "the phi" is abortive in that case. The semantic content of *T*, the corresponding type, is a function that assigns F to *t*.

The last two paragraphs are *virtually* indistinguishable. The only difference lies in the order in which the quantifiers are presented. The referentialist position gives wide-scope to the "for some x": it is the first quantifier in the sentence that gives his analysis. The Russellian position gives that quantifier much narrower-scope; it occurs within the scope of several other quantifiers. Because of this difference, the referentialist has "… the inventor of bifocals…" be *de re* about Franklin, whereas the Russellian has such sentences *not* be de re about him. But notice how slight the difference is between these two analyses: this corresponds to the fact that, contrary to what is commonly thought, the referentialist position *is* consistent with the cognitive significances of sentences containing definite descriptions.

The character of a sentence-type containing an indexical is not a function from contexts to propositions. The semantic content of "I am tired" is a function that assigns meaning to its tokens. If x uniquely utters a token *t* of "I am tired", the meaning of *t* is a function F that assigns success to *t* exactly if x is tired. The semantic content of the type "I am tired" is a function that assigns F to *t*.

The semantics just outlined is obviously Montague-inspired. But it makes some changes to Montague's semantics, corresponding to the developments in semantics that have occurred since the time of his writing. Our analysis accommodates the fact that tokens of proper names and indexicals are directly referential. It also accommodates the fact that expression-types have a very different semantic content from expression-

tokens, even where proper names, and other apparently one-dimensional expressions, are concerned.

Chapter 16 Truth as Instantiatedness and the Unity of the Proposition

In any discussion of semantics, one encounters innumerable references to "propositions." But no satisfactory account of what propositions has been given. Some philosophers (e.g. Quine) deny the existence of propositions. The grounds for this are usually that propositions, if they exist, are platonic entities, and that platonic entities don't exist. But attempts to do without propositions have not faired very well. Further, the belief that, in our analyses, we ought to *try* to do without propositions rests, I believe, on a number of misunderstandings. Given a correct understanding of what propositions are, there turn out to be no reasons to deny their existence or to otherwise regard them with suspicion, and they turn out to be explanatorily powerful and well-behaved entities. Further, such an analysis gives us fresh and deep insights into hitherto stagnant quarters of semantics and (I hope to show in the next volume of this work) into psychology.

The purpose of this chapter is to say what, in my view, propositions are, and to make corresponding adjustments to a number existing semantic views.

Consider the proposition

(*) *there is a cactus in front of my window*

A world where (*) is true is different from one where (*) is false. In the one kind of world, there is an instance of the property of being a cactus in a certain place, at a certain time. In the other kind of world, there is no instance of that property. It thus seems that for (*) to be true is for various properties to be instantiated and that for (*) to be false is for those same properties *not* to be instantiated.

As it happens, (*) is false. But I can easily enough imagine what things would be like if it were true. I imagine a world where, at certain places

and times, there are instances of certain properties.

What is true of (*) is obviously true of other propositions. These reflections suggest a certain view as to what propositions are, and as to what it is for a proposition to be true. A proposition is a collection of properties, and a proposition is true if the relevant properties are instantiated.

I thus propose that a proposition is a set of properties, and that for a proposition to be *true* is for all of those properties to be instantiated. Truth is identical with instantiatedness.

In this paper, I will develop the suggestions just put forth. There are atomic propositions and molecular propositions. The position in question can be made to apply to both.

§ There are two requirements that any adequate theory of propositions must satisfy. First, it must accommodate the fact that a propositions has a certain *composition* – or, as Katz put it, a certain *de*composition. Second – this is related – it must account for the *unity of the proposition*; it must explain how a proposition's various constituents can combine into something true or false, instead of remaining a truth-valueless set or heap.

The proposition *Socrates punched Plato* has a unique decomposition into minimal parts; and that proposition is obviously identical with a certain arrangement of those parts. The question is: *how* exactly are those parts arranged in that proposition? What conceivable arrangement of Plato, Socrates, and the relation of punching constitute a true proposition? The set (Plato, Socrates, the relation of punching) obviously isn't true or false. It doesn't matter whether we impose some kind of ordinal structure on that set. The *ordered* triple <Plato, the relation of punching, Socrates> isn't true or false. Neither is <the relation of punching <Plato, Socrates>>, or any other ordering of those three constituents. So even though it seems clear that *Plato punched Socrates* is identical with *some* arrangement of those three things, it is far from clear what kind of arrangement is the appropriate one. In my view, it becomes quite clear what the right kind of arrangement is, so long as we pursue the suggestions made earlier regarding the nature of propositions.

(A) "Smith punches Jones"

is a sentence.

(B) "Smith, the relation of punching, Jones"

is not.

A sentence is not just a heap of referring terms. A corresponding point holds about propositions.

(Ap) *Smith punches Jones*

is a proposition.

(Bp) *Smith, the relation of punching, Jones*

is not. A proposition is not a heap of objects and properties. A proposition is a set of objects and properties that are *unified* in some way. But *how* must they be unified? In what does the "unity of the proposition" lie? This question has never been adequately answered. I believe that our analysis of propositions gives us an answer to this question.

§ Frege's answer was to say that, in a proposition, certain items are "unsaturated": they have holes in them. These holes enable the items in the proposition to link together in the right way. In (A), the expression "punches" corresponds to something with holes in it; these holes enable Smith and Jones to link up with that thing in the appropriate way. By contrast, in (B), the thing corresponding to "the relation of punching" doesn't have the right number of holes in it; so it doesn't appropriately link up with Bob and Smith.

Frege's "solution" is really just a way of labeling the problem, as Davidson said.[1] Obviously we aren't talking about physical holes. We are talking about holes in a metaphorical sense. What we want to know – what Frege doesn't tell us – is what this metaphor is a metaphor for.

The difference between (A) and (B) obviously has something to do with the fact that the first has grammar, whereas the second does not. In the first, there are inflections. (I will argue later some of these are not phonetically apparent.) In the second, nothing is inflected. The difference between a proposition and a non-propositional heap is somehow coded in these grammatical articulations. The question is: what exactly these articulations do? Do they themselves refer to entities of some kind? Does the "-s" on "punches" refer to some mysterious entity – perhaps to kind of function? Is *that* why (A)

encodes a proposition while (B) does not? Are grammatical articulations *referring terms* with the special property that the things to which they refer confer unity on what would otherwise an unordered heap? Or do grammatical articulations not refer to anything? Do they exert their unifying influence in some other way?

§ Let me start out by stating, in extremely rough outline, the position I will defend. Consider the ordered-set <2,3>. The ordinal properties of this set can be coded in, and actually identified with, the properties of some *unordered* set, namely ((2), (2,3)). I will argue that propositions are unordered sets of properties. Thus *Smith punches Jones* is a set of properties. For a proposition to be *true* is specifically for all of its constituents to be *instantiated.* Truth is instantiation. For a proposition to be true is for all the members of a certain kind of set to be instantiated.

Here two worries immediately arise. First, it seems obvious that *Smith punches Jones* and *Jones punches Smith* have the same *constituents*, even though they are obviously different propositions. I will argue that those two propositions *share* constituents – they have in common Jones, Smith, and the relation of punching. But the one proposition has constituents not had by the others; and the ordinal differences between those two propositions are reflected in these differences in their constituents.

The other worry is this: Surely Smith is not a *property*; surely Smith cannot be instantiated. Let me outline what I will say about this. On my analysis, (Ap) is identical with some set S that comprises only things that can be instantiated, and that therefore comprises only *properties*. I will argue that what is a member of S is not Smith *per se* but is rather the property of being identical with Smith. Obviously *that* thing can be instantiated.

No doubt, this maneuver will raise another concern: a concern I will spend a lot of time trying to allay. It will at once be said that what is a constituent of (Ap) is *Smith* and not the property of being identical with Smith. (Ap) says that *Smith* punched Jones, not that the property of being identical with Smith did so.

I fully agree: what is a constituent of (Ap) is Smith, and Smith – nothing else – is what punched Jones. But we need to be very clear what it means to say that something is a *constituent* of a proposition. As I will argue, when you say that Smith is a "part" or "constituent" of a proposition P, you are not saying that Smith's relation to P is anything like a brick's relation to a

house or a wheel's relation to a car. What you are saying is that the truth of P depends in some way (that we will specify) on Smith: the "constituency" relation in question is a relation of *dependence* – it is a purely logical relation, and cannot be modeled on a spatial relation of any kind. Now suppose we identify (Ap) with some set S, one of whose members is the property of being identical with Smith. By doing so, I will argue, we can do justice to the relation that holds between (Ap) and Smith – the relation of dependency that holds between them. So far as it is meaningful, let alone true, to say that Smith is a "part" or "constituent" of a proposition, what is being said is precisely that there exists just the kind of dependency relation we've been discussing. Our construal of propositions as sets of properties, along with our construal of truth as the instantiatedness of the members of such sets, captures that dependency relation to a tee.

There is more to say on this matter. Consider the following two statements:

(i) x is a constituent of a proposition P.
(ii) y is a member of set S.

The sense in which x is a part of P is entirely different from the sense in which y is part of S. x is a part of P only in the sense that the truth of depends on some fact about x. y is a part of S in some other sense. It is true that I identify

(Ap) *Smith punches Jones*

with some set S; it is also true that the property *identical with Smith* is a member of S, and not Smith himself. But the members of S are not the constituents of P. For Smith to be a constituent of (Ap) is specifically for (Ap) to be about Smith -- it is for there to be a dependence relation between Smith and (Ap). Precisely that dependence-relation holds between Smith and (Ap) if we identify (Ap) with S, and then go on the truth of (Ap) with the being-instantiated of all the members of S. So our analysis *is* consistent with the truth of statement that Smith is "constituent" or "part" (Ap), at least on any significant disambiguation of that statement. It is irrelevant that Smith himself is not a member of S. When we are talking about the membership of S we are talking about one thing; when are talking about the constituency of P we are

talking about another (we are taking about relations of dependency or aboutness). There is a *connection* between these two sets of facts: facts about the constituency of P – facts about the aboutness relations in question – are *realized by* or *supervene on* facts about the membership of S. But, strictly speaking, the two sets of facts are distinct. We will explore this further in due course.

§ There is yet another worry. Some may object to the idea that there is such a thing as the property of being identical with Smith. It will be said that tallness is a property, and so is wetness – but there is no such thing as the *property* of being identical with Smith.

I think this point of view is wrong for two different reasons. First, statements like "so and so is not identical with Smith" are obviously meaningful – indeed, they are often true -- and it is hard to see how this could be the case unless there were such a thing as being identical with Smith.

Second, suppose that Smith exists in some world w. Obviously Smith is not a *basic* constituent of w; he is not one of its lower level constituents – he is not a quark or a muon or a single displacement of mass-energy. All of Smith's activities – indeed, his very existence – supervene on lower-level states of affairs. So *if* Smith occupies some part of space-time, that supervenes on much more basic facts – facts about the displacements of mass-energy (or whatever the basic constituents of the world in question are). Nobody is going to deny the existence, let alone the scientific respectability, of the properties involved in those lower level states of affairs – charge, spin, mass, and so on. Nobody is going to say that there is anything suspect or "viciously circular" about believing in these properties. Every fact about Smith, even his very existence, is fixed by facts about those wholly acceptable, lower-level properties. So the property of comprising Smith is a perfectly respectable property of a world. It is a necessary consequence of there being certain patterns of mass-energy displacement in that world. Given the respectability of these lower-level states of affairs, and given that Smith's existence is a necessary consequence of them, it is hard to see the property of being Smith in any way lacks respectability – it is hard to see why we should deny existence to the property of being identical with Smith.

We might even think of it this way. It is clear that aesthetic and moral properties are realized by lower-level states of affairs: certain patterns of mass-energy displacements realize beautiful works of art, while others don't. Given that Smith's existence is realized by lower-level states of affairs -- by certain

patterns of mass-energy – we might even think of Smith as a kind of *property*. If a pattern of mass-energy displacements has the right properties – if, perhaps, it has the right origins, and it has the right internal properties – then it "Smiths".

Of course, there are profound differences between the property of being beautiful and the property of Smithing. A world can comprise many beautiful things; but – leaving aside things like cloning, meiosis, and so on – a world can comprise only one Smith. The reason for this is not hard to identify. Whether a pattern of mass-energy displacements is beautiful doesn't have anything to with its having certain origins – the Mona Lisa would be beautiful in any place or time. By contrast, whether such a pattern Smiths or not is tied very closely to its conditions of origination. A qualitatively identical pattern that had different origins wouldn't be Smith – it would realize a Smith-duplicate, not Smith himself. But here the important point is that, quite plainly, whatever is the case with Smith – indeed, whether he exists or not – is fixed entirely by scientifically respectable lower-level properties. So it hard to see why we cannot countenance the property of realizing Smith or, therefore, of being identical with Smith.

The analysis given here applies to analytic as well as synthetic propositions, and atomic as well as molecular propositions.

§ I will argue that grammatical articulations *are* what distinguish word-heaps from sentences. I will also argue that grammatical articulations do not typically refer to anything. (In *some* cases they may, but not typically.) If grammatical articulations referred, then they would simply create big heaps where, previously, there were only small heaps. Suppose the inflections in (A) refers to entities x, y, and z. In that case, (A) wouldn't be a proposition; it would be the heap: *Smith, Jones, the relation of punching, x, y, and z.* Referring to new things doesn't turn a heap into a proposition; so grammar articulations couldn't perform their unifying function by being referring terms. What grammatical articulations do is to indicate how to generate new properties on the basis of the properties introduced by the referring terms. (I am treating relations as properties – an n-place relation is a property had by n-tuples.) Consider (A). The referring terms – "Smith", "punches", "Jones" – pick out Smith, the relation of punching, and Smith. (For reasons I will give, Smith and Jones are themselves properties.) The purely *grammatical* properties of that sentence – the inflection on the verb, the relative positions of

"Smith" and "Jones" – indicate how *new* properties are to be generated from those primitive properties. Those new properties – we will see soon enough what they are – suffice to encode the ordinal differences between *Smith punches Jones* and *Jones punches Smith.*

II. There are, ultimately, three conceptions as to what propositions are.

(i) Functions from worlds (or situations) to truth-values. So *that Plato snores* is the set of worlds where Plato snores or is a function that assigns truth to worlds where Plato snores, and falsity (or no truth-value) to other worlds.

(ii) States of affairs. So the proposition *that Plato snores* is identical with the event of Plato's snoring (or, perhaps, with a series of such events).

(iii) Arrangements of objects and properties. So *that Smith punches Jones* is some kind of arrangement of Smith, Jones, and the relation of punching, much as <2,6,8> is a structure consisting of three numbers

I believe that (i) and (ii) are false. I believe that (iii) is closer to the truth. But, first, I don't think it is quite correct. (As I will argue, *that Smith punches Jones* is not an arrangement involving Smith and Jones *per se*, but the properties of being identical with Smith and Jones, respectively.) But leaving aside this last point, (ii) is woefully incomplete. Suppose propositions *are* arrangements of individuals and properties. We still need to know two things. First, what is the relevant kind of arranging? We aren't talking about spatial arrangement. So what are we talking about? Second, how can an arrangement of objects and properties be *true or false*? The set (Smith, the relation of punching, Jones) isn't true or false? It doesn't matter how much ordinal structure we introduce: the ordered triple <Smith, the relation of punching, Jones> is no more true or false than its counterpart. We will deal with these problems in due course.

§ The Tractarian Answer

If I am reading him correctly, Wittgenstein gave a clear answer to the two questions we just asked. The sentence "Smith punches Jones" encodes an arrangement of objects and properties: Smith is related to Jones in some way,

and the relation in question is of that of puncher to person punched. If reality comprises a state of affairs in which those things are arranged in that way, then the sentence is true. If not, it is false.

Suppose this view is right. In that case, the proposition *that Smith punches Jones* consists of Smith, Jones, and the relation of punching being arranged in a certain way; and they are arranged in the very way in which they would be arranged in the corresponding state of affairs. The sentence is isomorphic with the proposition. The proposition is isomorphic with state of affairs. That is why the sentence represents the state of affairs.

There is an obvious problem with this. If this view is right, then the thing meant by "Smith punches Jones" consists of Smith, Jones, and the relation of punching being arranged in the very way they are arranged in the corresponding state of affairs. How are they arranged in that state of affairs? Smith is punching Jones. So if the view in question is right, the proposition *that Smith punches Jones* consists of Smith standing in the relation of punching with respect to Jones.

But this means that the mere existence of the proposition guarantees its truth. If the proposition *that Smith punches Jones* consist of Smith's punching Jones, then the mere existence of that proposition guarantees that Smith is punching Jones. But the proposition *Smith punches Jones* exists regardless of whether Smith punches Jones. It is obvious that the way in which Smith, Jones, and punching are arranged in the proposition cannot be identical with the way they are arranged in the corresponding state of affairs.

At the same time, it does seem fair to say that *that Smith punches Jones* consists of Smith, Jones, and punching in some kind of arrangement. But, it would appear, the "picture-theory" seems not to give the right arrangement.

§ Are Propositions States of Affairs?

This naturally leads us to position (i): a proposition is a state of affairs. This is not a popular position, since there are two obvious problems with it. First, propositions are *true* or false. But a state of affairs is not true or false, and therefore isn't a proposition.[2] Second, (i) makes it impossible for there to be false propositions, unless one takes the heroic of measure of saying that there exist non-existent states of affairs.

Some (e.g. Barwise and Perry) have identified propositions with

possible states of affairs or, at any rate, with sets of such things. This is not so different from position (iii). Before we consider (iii), we should note one other reason why propositions cannot be states of affairs: this other reason is crucial to having a correct conception of what propositions are; it is also crucial to evaluating the Barwise-Perry position.

Consider (Ap). That proposition has a unique decomposition into minimal parts. It neatly decomposes into Smith, the relation of punching, and Jones. (It may have other ingredients, corresponding to tense-markers and the like. Let us ignore these for now.) Notice that, in that proposition, the relation of punching is completely distinct from Jones and from Smith. The proposition *connects* them, of course. But each of these things occurs in its purity: what occurs is not Jones' being tall or angry, but simply *Jones*. The proposition consists of three, neatly separated (though linked) entities.

In light of this, suppose that (Ap) is true. Consider the *state of affairs* that makes it true – imagine Smith punching Jones. Let S be that state of affairs. S doesn't have a structure even remotely like that of (Ap). S's minimal parts are definitely *not* the relation of punching, Smith, or Jones. First of all, Smith's act of punching is "adjectival" on Smith. It isn't as though Smith is here, and his act of punching is there. Smith's punching Jones is inseparable from Smith and Jones. Further, that act is by no means a *minimal* part of the states of affairs involved. It probably involved a number of subsidiary acts. (Even if it didn't, that doesn't affect our point. Consider the proposition *Beethoven composed the 9th symphony*. The "act" of composing involved innumerable deliberate subsidiary acts.) The same is true *mutatis mutandis* of Smith and Jones. They are not *minimal* parts of anything. They are constituted by innumerable events. If the state of affairs in question has minimal parts, those would be sub-atomic of some kind. But, I think, such events wouldn't really be parts of that event in quite the sense in which Jones and Smith are parts of the proposition *that Smith punches Jones.* It seems more natural to say that the state of affairs, unlike the proposition, doesn't have *any* unique, correct decomposition. In any case, it doesn't have a decomposition even remotely like that of the corresponding proposition.

Also, Smith and Jones are not encountered *in their purity* in the relevant state of affairs. What occurs in the proposition appears to be simply *Jones.* Of course, the proposition goes on to ascribe a property to Jones. But Jones' occurrence in that proposition is unpropertied. What occurs is not Jones' being tall or Jones' being sweaty - just *Jones.* But in any state of affairs,

even one that involves Jones' floating in a void, what occurs is not just *Jones*: it is a state of affairs involving Jones. What occurs in reality is Jones' being a in a certain place, having a certain height and weight, animated by certain attitudes or beliefs. There is no Jones *simpliciter.* There are states of affairs involving Jones. Jones is inseparable from them. It is not as though there is *first* a Jones and *then* a state of affairs is constructed out of Jones. For Jones to exist *is* for a state of affairs of some kind to exist. Jones is not accurately thought of as an isolable constituent of states of affairs. Jones *simpliciter* – the Jones that doesn't occur in a state of affairs – is a pure abstraction. What occurs in reality is not Jones *simpliciter;* what occurs are Jonesy states of affairs. There can no more be "free" occurrences of Jones – occurrences of Jones in isolation of a state of affairs -- then there could be free occurrences of redness. For these reasons, I think that *metaphysically* Jones is a property of states of affairs. As for what exact kind of property he is, that depends on your other metaphysical views. If you believe the essentiality of origins, then a state of affairs is Jonesy if it has a certain history. If you have a more theologically inclined view, you might say that a state of affairs is Jonesy if it is animated by the right kind of immortal soul. All that is important here is that Jones must be seen as a property of states of affairs, lest we run into insuperable metaphysical problems.

§ It is clear that the *articulations* one finds in a proposition are radically different from those found in the corresponding states of affairs. Ian Hacking refers to this as the "articulation problem". In the proposition *Socrates is tall*, we have one constituent corresponding to Socrates and a totally different one corresponding to tallness. But Socrates cannot be separated from his tallness. Even if you cut off Socrates' legs, you are not separating him from his height: you are only giving him a new height, and he will be as inseparable from his new height as he was from the old one. Propositions separate out what, in reality, are balled together. Propositions and states of affairs have entirely different decompositions. Various thoughts and atoms are constituents of Socrates. But no thoughts or atoms are constituents of the proposition *that Socrates is tall.* So propositions are not states of affairs, and they are not things that are even remotely *like* states of affairs.

However we analyze propositions, we must make sure that the things with which we identify propositions have the same decomposition as the propositions themselves. It is practically a datum that Socrates is not only a

constituent of the proposition *that Socrates is tall*, but a *minimal* and *discrete* part of it. The idea that propositions are states of affairs is inconsistent with that idea. For given any state of affairs in which Socrates is involved, he is absolutely not a minimal part of it, and he is absolutely not neatly separable from it. This fact, I believe, vitiates the idea that a proposition is a set of states of affairs or, what is not so different, a set of worlds. A set of states of affairs has whole states of affairs as ultimate constituents: Socrates himself is no ultimate constituent of such a set. And any given state of affairs cannot possibly have Socrates as an ultimate constituent.

§ This brings us to approach (iii) Consider the set of worlds where Socrates is tall. Socrates isn't a constituent of that set in remotely the sense in which, by all appearances, he is a constituent of *that Socrates is tall*. Now consider a function that assigns truth to each of the worlds just mentioned. The point we just made still applies. On face of it, Socrates isn't a constituent of that function in the sense in which he is a constituent of *that Socrates is tall.*

Of course, attempts have been made to reintroduce constituent-structure into that function – to make it more like the corresponding propositions. Whether these attempts succeed is a question for another time. Even if they do, the possible-worlds approach is still a failure. If a possible world is a set of propositions, then (ii) is viciously circular. If a possible world is a concrete alternative world, then (ii) is committed to all manner of dubious metaphysical and epistemological views. (There is no point in going through them here: they are well documented.) Given this, it would probably be better to abandon (ii).

§ A word of clarification is in order here. I think that it is *true* to say that propositions are functions from worlds to truth-values. But such a statement is unacceptable *if taken as an analysis*. There are true, and even informative, statements that cannot serve as analyses. Permit me to explain my meaning. First of all, there is surely nothing wrong with talking about hypothetical worlds – such talk is innocent. Let w be a hypothetical world. Some propositions will be true in w and some won't. Let P be a proposition that is true in w. Since P is true in w, we can naturally think of P as associating w with the truth-value *true* (or the property of truth). Let w* be some other hypothetical world, but suppose that P is false in w. P associates falsity with w*. And so on, for any other hypothetical world one might consider. So it is

certainly true that propositions associate truth-values with hypothetical worlds, and can thus be regarded as functions from worlds to truth-values.

But propositions *are* functions from worlds to truth-values in the sense of "are" that indicates *predication.* All and only propositions have the property of assigning truth-values to worlds. But if we *analyze* propositions as functions from worlds to truth-values , then our statement is either viciously circular (if the worlds in question are treated as propositions) or is wrong for other reasons (if the worlds are treated as concrete objects).

A comparison might be in order. It may well be *true* that x is the same event as y iff x and y stand in the same causal relations, i.e. if everything that causes x also causes y and if anything that x causes is also caused by y. And such a statement is not only (quite possibly) true, but also informative. But taken as an *analysis* of the concept of an event, that statement is circular: after all, causes and effects ere themselves events.[3]

Here is another example. Nothing that is not a proposition entails anything; and anything that is a proposition entails something. So

(*) all and only propositions stand in entailment relations.[4]

is true. Further (*) is an *informative* statements. But it is not acceptable as an *analysis* of the concept of a proposition. Anything entailed by a proposition is itself a proposition. So (*) amounts to:

(**) all and only propositions entail other propositions.

Taken as an analysis, (**) is viciously circular. I believe the statement *propositions are functions from worlds to truth-values* is informative and accurate. But it is viciously circular if it is taken as an analysis. I believe that PWS is a giant projection of a failure to distinguish the conditions that must be met to provide an analysis of a concept from those that must be met to make a true statement about a concept.

(i) is an out and out failure. (ii) is either viciously circular or is extremely dubious. Further, it is at best an open question whether a set of worlds or a function from worlds to truth-values could have a decomposition at all like that of a proposition. Given this, we should, I believe, try to make some variant of (iii) work.

III. Now it is time to give a positive solution. Let us start with some basic points about propositions. To a first approximation, a proposition is a property of a world. Suppose that *Plato snores* is true in w and false in w*. In that case, w has a property that w* does not. w comprises certain states of affairs that w* does not. The space-time manifold in w has a ripple in it *not* had by the space-time manifold in w*. Kenneth Taylor (1998) speaks of states of affairs as "ripples in the quantum". For Socrates to be snoring is for the quantum to be rippled or indented in a certain way (with the possible qualification that those ripples have to have certain origins).

Being tall or red is a way that individuals can be. Thus tallness and redness are properties of individuals. A proposition is a way a world can be. It is thus a property of a world. This is why David Lewis (1986) explicitly identifies propositions with properties of worlds. For the proposition *Socrates snores* to be true in w is for w to be a certain way – it is for w's quantum to have certain ripples in it. That proposition is thus a property, and that proposition is true in a world exactly if that proposition is *instantiated* in that world. We can thus identify truth with instantiatedness. A proposition P is a property of a world, and P is true exactly if that property is instantiated.

Here we should deal with what would appear to be a major problem with what we've said so far (to avoid confusion, I'm going to put the objection in the mouth of an imaginary interlocutor):

Let p be the property of being a world where

(*) Plato punched Socrates

is true. Let w be a world where the proposition Plato punched Socrates is true, and let w be a world where that same proposition is false. w has p. w* does not. So it is obvious that, for some property x, Plato punched Socrates is true in a world exactly if that world has x. This point generalizes without limit: given any proposition P, there is some property P* such that P is true in a world exactly if that world has P*.*

But here is the problem. The proposition seems to be more basic than the corresponding property: the property must be understood in terms of the property; and the property's being instantiated must be analyzed in terms of the proposition's being true. This spells the doom of your analysis: for you are trying to analyze the proposition in terms of the property, and the

proposition's being true in terms of the property's being instantiated.

For (*) to be true in w, what property must (*) have? It must have p. But what is p? It is the property of being a world such that, in that world, (*) is true. So (*) is more basic than p. p is defined in terms of (*). This means that we cannot analyze (*) in terms of p; and we cannot analyze what it is for (*) to be true in terms of p's being instantiated. On the contrary, we must analyze p in terms of (*), and we must analyze p's being instantiated in terms of (*)'s being true.

This point generalizes without limit: what we just said about (*) is true of any proposition. Consider the proposition

(**) penguins don't fly.

Obviously there is some property P such that (**) is true in a world w exactly if w has P. P is the property of being such that (**) is true in it. But we cannot analyze (**) in terms of P, and we cannot analyze (**)'s being true in terms of P's being instantiated. On the contrary, we must analyze P in terms of (**), and P's being instantiated in terms of (**)'s being true. So your account seems doomed to vicious circularity.

I think that this point is entirely wrongheaded. (To simplify discussion, let us leave momentarily aside analytic propositions.) A proposition is never *just* true. It is always true in virtue of something about the world: mass-energy is distributed in the right way; sentient beings are having the right thoughts or feelings. A true proposition registers some fact, and the truth of a proposition supervenes on something more basic.

Consider (*). It is not as though Plato *just* punches Socrates. Plato's punching Socrates supervenes on facts of a much more fundamental kind: Plato punches Socrates in virtue of the fact that atom a moves in this way at this time, atom b moves that way at that time, and so on. Given two worlds where the same elementary particles move in the same way, either (*) is true in both of them or in neither. So the truth of (*) supervenes on the way in which mass-energy is distributed in the world. That distribution could obviously be described without mentioning Plato or Socrates or the relation of punching. Given any one of the electron-jumps involved in Plato's punching Socrates, that electron-jump could obviously be described without mentioning Plato or

Socrates. The same applies to each of the micro-events involved Plato's punching Socrates. Obviously the totality of those micro-events could (in principle) be described without mentioning Plato, Socrates, or the relation of punching. That totality is plainly *more* basic than (*). If (*) is true in a world w, that fact supervenes on the way in which space-time is wrinkled in w. Those wrinkles *underlie* the truth of (*); and they *can* plainly be described without mentioning (*). Thus the property of having those wrinkles is more basic than (*), and the property of having those wrinkles is more basic than the concept of (*)'s being true. So there is no circularity in analyzing (*) in terms of the property a world has if it is wrinkled a certain way, and there is no circularity in analyzing (*)'s being true in terms of a world's having the right wrinkles.

There is more to say in this connection. As we've noted, given any two worlds where mass-energy is distributed in the same way – any two worlds whose quanta are rippled in the same way – (*) is true in both of them or in neither: so the truth of (*) supervenes on other facts. But the converse is not true. There can be two worlds where (*) is true but that otherwise differ enormously. There are many different ways that Plato can punch Socrates, and there are many different ways that the surrounding circumstances can vary. So whether (*) is true is fixed by how the quantum is rippled in w. But how the quantum is rippled in w is by no means fixed by whether (*) is true in that world. This puts it beyond doubt that, if (*) is true in a world w, that is in virtue of w's having some property that is *more basic* than (*): it is in virtue of facts about the basic constituents of w – facts about the way the quantum is rippled in w. The property that a world has if (*) is true in it is thus less basic than the property that a world has if the quantum is rippled the right way in it. So, contrary to what the objector says, there is some property P such that P is more basic than (*), P's being instantiated is more basic than (*)'s being true, and such that (*)'s being true is fixed by P's being instantiated (but not *vice versa*). Thus our account is *not* guilty of vicious circularity, despite what the objector says.

These points generalize. Consider any empirical proposition p. p is not *just* going to be true in w. It is going to be true in virtue of some fact about the basic constituents of w. p will be true in w in virtue of the fact that, as we put it, the quantum in w is rippled a certain way. The property of having that rippling will be *more* basic than p, and a world's having the right rippling will be *more* basic than p's being true. So I think the objector's point is misguided.

Let us sum up. Consider the proposition *Jones punches Smith.* This is true in w iff the "quantum" in w is "rippled" a certain way – if w comprises displacements of mass-energy of a certain kind. So the proposition *that Jones punches Smith* can be thought of as a property of worlds, and that proposition/property is true in a world w iff it is instantiated in w.

IV. But the story just told isn't quite accurate. There are a couple of reasons for this. Let p be the property that a world must have if, in it, Jones punches Smith. p will be instantiated in w exactly if, in w, there is a certain pattern of mass-energy displacements. But exactly these mass-energy displacements will suffice for the truth of infinitely many other propositions. Obviously they will suffice for the truth of every analytic proposition. (Since such propositions are true under any circumstances, it follows vacuously that whatever happens in a world is sufficient for their truth.) But those same displacements will make true *somebody punches Smith* and *Jones punches somebody* and *somebody punches somebody* and *somebody does something intentionally.*

We could try to get around this. If a mass-energy displacement is enough to make true *Smith punches Jones*, it is enough to make true *somebody punches Jones*, but the converse doesn't hold. So perhaps we could use this asymmetry to distinguish the kinds of properties associated with the patterns of mass-energy displacements.

But this maneuver won't always work. Consider the proposition *x is a circle* and *x is a closed two-dimensional figure of uniform curvature.* Surely these are different propositions; and surely in any possible world, any mass-energy displacement sufficient for the truth of the one is sufficient for that of the other. Of course, some will deny that these are different propositions. But that denial has little independent motivation – it seems an ad hoc way of ameliorating a commitment to a degenerating research program.

Also, the Lewisian picture seems to make analytic propositions disappear. If a proposition is a property of a world, then what a proposition tells you is nothing more and nothing less than how a world with that property is different from a world without it. An analytic proposition is one that all worlds have. So, on Lewis' model, analytic propositions don't tell you anything: they might as well not exist.

The problem with the view we've been considering is one we've already encountered. Propositions have decompositional structure. A

proposition is not specified merely by its truth-conditions. It is necessary also to give its decomposition The Lewisian analysis just considered doesn't accommodate the internal or decompositional differences between

(i) *Alpha is a figure whose perimeter coincides with the class of all points equidistant from a given point in a plane*

and

(ii) *Alpha is a closed two-dimensional figure of uniform curvature.*

So it is not enough to say that a proposition is a property of a world. This obviously relates to the fact that analytically equivalent propositions can convey very different information and, what is a consequence, that analytic propositions *do* typically convey information. At the same time, the view just considered contains some profound truths, and our view will be more a development than a rejection of it.

Another illustration may help. Again consider:

(Ap) *Smith punches Jones.*

In this world, if Jones and Smith are to exist, innumerable sub-atomic, atomic, chemical, molecular, cellular…events must occur. In *any* world where Jones and Smith exist, their existence supervenes on the existence of lower-level phenomena. The same is true of the relation of punching: if there is a punching, that supervenes on many other events. So no state of affairs that makes (Ap) be true has Smith or Jones or the relation of punching among its ultimate constituents. Given this, suppose we identify (Ap) with some property P such that (Ap) is true in w exactly if P is instantiated in w. In any world, an instance of P will make true many propositions whose ultimate constituents are *not* Smith or Jones – whose ultimate constituents are tiny displacements of mass-energy. This makes it hard to believe that P itself has a unique decomposition whose ultimate constituents are specifically Smith, Jones, and the relation of punching. If Jones manages to punch Smith, that is only by the grace of innumerable subvenient states of affairs. So P's being instantiated demands the occurrence of subvenient states of affairs having certain properties. So it is individuative of P, and thus essential to it, that whenever it

is instantiated, things *more basic than* Smith, Jones, and the relation of punching are instantiated. For this reason, it cannot be said with confidence that P has a unique, ultimate decomposition into Smith, Jones, and the relation of punching. On the contrary, it seems as though those three things are *not* among its ultimate constituents. This makes P unsuited for identification with (Ap).

All of this assumes that properties themselves even *have* constituents. It is clear that instances of properties have constituents. An instance of the property *chair* has legs and a back. But it isn't really clear if that property itself has constituents. Even if properties do have constituents, it isn't clear if they have them in the sense in which propositions have constituents. Perhaps the property *water* is composed of the property *hydrogen* and the property *oxygen.* But are the last two properties constituents of the first in the very sense in which *Smith* is a constituent of (Ap)? This is, at best, an open question. I will argue that the answer is "no".

Let us sum up. It is pretty clear that Smith and Jones and the relation of punching are constituents of (Ap). It is also clear that (Ap) has a unique decomposition into minimal parts, and that the three things just mentioned are among those minimal parts. But it is *not* clear if P has an even remotely comparable decomposition. It isn't even clear if P has *a* decomposition in the sense in which (Ap) does. Also, P's being instantiated, while sufficient for the truth of (i), is also sufficient for the truth of infinitely many distinct propositions. So it does not seem right to *identify* (i) with P.

§ There is one other desideratum that any theory of propositions must satisfy. Consider the following propositions:

(Ap) *Smith punches Jones.*
(2) *Smith punches somebody.*
(3) *Smith punches everybody.*
(4) somebody punches somebody.
(5) Fred punches nobody.
(6) Fred punches Jones.
(7) Nobody punches Jones.

It is pretty clear that all of (Ap)-(&) have a constituent in common, namely: *__punches__* or *x punches y.*

It is pretty clear that (Ap)-(3) that have a constituent *in common*, namely: *Smith punches__* or *Smith punches x.*

It is pretty clear that (Ap)-(3) that have a constituent *in common*, namely: *Smith punches__* or *Smith punches x.*

It is pretty clear that (6)-(7) that have a constituent *in common*, namely: *__punches Jones* or *x punches Jones.*

The naïve thing to say is that (Ap) has *three* constituents: one corresponding to Jones, one corresponding to Smith, and one corresponding to the relation of punching.

But that would overlook the fact that (Ap) has the component *x punches Jones* and also the component *Smith punches x.* One of (Ap)'s components seems to correspond to the property of being a thing x such that Smith punches x. Another seems to correspond to the property of being a thing x such that x punches Jones. So (Ap) has several constituents. It has one corresponding to Jones; one corresponding to Smith; one corresponding to the relation of punching; one corresponding to the property of being a thing x such that Smith punches x; one corresponding to the property of being a thing x such that x punches Jones; and, finally, one corresponding to the result of putting all of these things together.

Frege argued that the proposition *Smith punches Jones* is built out of the relation *x punches y*, and the functions *x punches Jones* and *Smith punches x.* The point we just made obviously has a certain similarity to Frege's. (In fact, I will argue, they are the same point.)

§ Here is how I would propose to accommodate all of these facts. (Ap) is a *set* S of properties, *one* of whose members is P. But P is not the only member of S. What are its other members? My answer to this question will initially be alarming.

Let us start with terminology:

Ps: the property of being identical with Smith.
Pj: the property of being identical with Jones.
Psx: the property of being a thing x such that Smith punches x.
Pxj: property of being a thing x such that x punches Jones.

Pp: the relation of punching.
P: as previously defined.

I propose that (Ap) is a set S whose members are exactly the properties just listed. And I propose that for (Ap) to be *true* is simply for all of its members to be instantiated.

For reasons already given, our including P on the list of relevant properties does *not* constitute a case of vicious circularity.

Let me now give the motivation for this analysis. First of all, under what circumstances is (Ap) true? The following properties must be instantiated: the relation of punching (i.e. the relation *x punched y*); the property of being a thing x such that Smith punched x; the property of being a thing x such that x punches Jones; the property of being identical with Smith; the property of being identical with Jones; and finally P.

If any one of these properties is not instantiated, then (Ap) is not true. For example, if the property *identical with Smith* is not instantiated, then there is no Smith; and there is therefore no Smith to punch Jones. So for (Ap) to be true, it is *necessary* that each of those six properties be instantiated.

It is also *sufficient.* Suppose that each of those properties is instantiated. In that case, there is a Smith, a Jones, a case of somebody's punching somebody, a case of Jones' being punched, a case of Smith's punching somebody, and a case of Smith's punching Jones. (The instantiating of P guarantees the occurrence of the last of these.)

There is more to say. The proposition (Ap) has discrete parts corresponding to Smith, Jones, the relation of punching, the property of being a thing x such that x punches Jones, and the property of being a thing x such that Smith punches x.; and (Ap) also has a (maximal) discrete part corresponding to the result of putting these together.

The same thing is true of S. S has one member corresponding to Smith, one corresponding to Jones, one to the relation of punching, and so on. So S has a breakdown parallel to that of (Ap).

So here is what we have thus far. The instantiating of those six properties is both necessary and sufficient for the truth of (Ap). Further, for each discrete constituent of (Ap), there is one property on that list; and for each property on that list, there is one constituent of (Ap).

Here we should address a worry one might have about our analysis?

The instantiating of P *by itself* is necessary and sufficient for the truth of (Ap). So why not just say that (Ap) is identical with P and that (Ap)'s being true is identical with P's being instantiated?

The reason is that P does not decompose in the same way as (Ap). If we are claiming that some thing x is identical with (Ap), we must at all costs make sure that x has Smith, Jones, and the relation of punching, and so on, as discrete constituents. If we identify (Ap) with P alone, it is by no means clear if that burden is met. But if we identify S with the set containing exactly the six properties listed above, then, first, (Ap) is true exactly if every member of S is instantiated and, second, the set just described has the same decomposition as (Ap): like (Ap) that set consists of Smith, Jones, the relation of punching, and so on.

As we said earlier, any analysis of propositions must accommodate the fact that different propositions can have common constituents. Our proposition has that virtue. Consider the proposition:

(PJ) *Plato punches Jones.*

Here is what we pre-theoretically believe about the relation between (Ap) and (PJ). They have certain constituents in common: one of these constituents corresponds to the relation of punching; another to the property *x punched Jones*; and another to Jones. For various reasons, I think we should hold onto this pretheoretic intuition. The facts relating to the "systematicity" and "compositionality" of thought accord very well with this intuition; they would not accord so well with its negation.

Now remember what we said about (Ap). Suppose that we apply exactly the same analysis (mutatis mutandis) to (PJ). What we end up with this is this. (PJ) is a set S* whose members are exactly the following properties:

Pp: the property of being identical with Plato
Pj: the property of being identical with Jones.
Ppx: the property of being a thing x such that Plato punches x.
Pxj: property of being a thing x such that x punches Jones.
Pp: the relation of punching.
P*: just like P (mutatis mutandis).

And (PJ) is true exactly if every member of S* is instantiated.

S and S* have certain members in common: Pj; Pxj; and Pp. These members corresponding to the constituents that we pre-theoretically believe (Ap) and (PJ) to have in common, namely: Jones; the property of punching Jones; and the relation of punching. So our analysis *accommodates* the fact that (Ap) and (PJ) have exactly those constituents in common. For exactly similar reasons, it *accommodate* the fact that both those propositions have one constituent in common with *Fred punches Barney*; and that one, but not the other, has a constituent in common with *Fred punches Jones.* For the same reason (mutatis mutandis), given *any* two propositions that we pre-theoretically believe to share some constituent c, our analysis is consistent with the fact that they share that constituent.

§ I wish to stress one point. We identified (Ap) with a set S containing all and only the six properties listed above. But it might seem that our inclusion of P on that list dooms our analysis to vicious circularity. P is the property of being a world where Smith punches Jones. So P is defined in terms of the proposition *Smith punches Jones.* But we are analyzing that very proposition in terms of P. So our analysis is circular.

But there is no circularity. We need some way to *identify* that property. And the easiest way to do so is in terms of a proposition – in terms of (Ap). When we *describe* an event, we typically use a sentence and, therefore, a proposition. (I say "typically" and not "always" because an event can also be described through some analogue medium, like a motion picture.) We say "Smith punched Jones" or whatnot. But consider the mass-energy displacements which realize that event; consider the over-all mass-energy storm constituted by those various displacements. That storm doesn't have remotely the kind of structure had by any of the propositions which describe it.

It isn't as though Jones can *just* exist. Jones' existence supervenes on that of innumerable other states of affairs. In fact, the entity *Jones* is really a kind of abstraction from these various states of affairs. In the proposition, *Jones* is separate from his various attributes. Jones isn't separate from his tallness, even though the proposition *Jones is tall* has one constituent corresponding to Jones and a discrete one corresponding to his tallness. The same is true (mutatis mutandis) of Jones' being punched Smith. the proposition *Smith punches Jones* has (at least) three discrete constituents. But the event thereby described – the storm of mass-energy displacements -- is a seamless whole; in any case, so far as it does have articulations, they are not

remotely like those of that proposition.

There are, in principle, different ways to identify that storm. It can be indicated by ostension. In fact, that is how it *was* identified by those who witnessed it. ("Harry, did you see *that*? Wasn't *that* astonishing? I never thought Smith was such a cad.?") Of course, that event can also be indicated by description. When we *describe* an event, we typically use a sentence and, therefore, a proposition. (We don't have to use propositions: we could also use analogue media like motion-pictures or painting.) Under the circumstances, it was necessary for us to *identify* that event by way of a sentence and, therefore, a proposition. But we didn't *have* to so identify it. In principle, it could have been identified through an ostension or a drawing or a motion-picture.

Now consider the *kind of* mass-energy storm we've just been discussing. In any world where (Ap) is true, there is an instance of that kind. To *identify* that kind, we use a sentence and, therefore, a proposition: we say, for example, "it is the way of being wrinkled that is common to all and only those worlds where *Smith punched Jones* is true." But we are using that proposition "referentially", not "attributively"; we are using it as a means of identification only. We are analyzing (Ap) in terms of the corresponding *kind* – not in terms of the proposition used to *pick out* that kind. We use the proposition *only* as a means of identification. It is the thing identified, not the means of identifying it, that enters into our analysis. So our analysis is not circular.

But there is a possible objection to this:

You say "consider the kind instantiated by all and only the mass-energy storms realizing the event of Smith's punching Jones." But that storm instantiates all manner of different kinds. Qualitatively different sorts of storms can realize a case of Smith's punching Jones. Smith and Jones can be located in any number of places and any number of times. The punch can be thrown in all kinds of different ways. And so on. So when you talk about the kind of mass-energy storm A given storm instantiates all manner of different kinds. So which kind are you talking about? You are talking about the kind instantiated by all and only those storms sufficient for the truth of Smith punched Jones. So that proposition is not merely a way of picking out a property; it enters into the very definition of the right kind of property; so your analysis is circular.

(Ap) is a contingent proposition. In this context, let us focus on such propositions. A proposition doesn't *just* hold. It holds in virtue of some fact – some state of affairs, some mass-energy displacement (or, more likely, consisting of such displacements). If a proposition is true in a world, there is something that makes it true – something that anchors it. Propositions aren't free-wheeling. A true proposition is anchored to the world by a non-propositional entity.

There are a couple of corollaries. First, if a proposition is true in a world, that imposes constraints on the facts constituting that world. Given that *Smith punched Jones* is true in w, there are some limits on how the underlying facts can be. There are some limits on what the underlying mass-energy displacements can do.

A consequence is that if a given proposition is true in two worlds, then those worlds have something in common *in virtue of which* that proposition is true in both of them. It is not a *brute* fact that they are both true; the commonality is anchored in some extra-propositional commonality. If we say that there is no such commonality, we are saying that the underlying facts can be vary without limit; we are saying that the truth of that proposition imposes *no* limits on the underlying facts. The mass-energy displacements can occur in any way at all. But if a proposition is true, that plainly *does* impose limits on these underlying facts.

So if *Smith punched Jones* is true in two different worlds, that is in virtue of some *extra*-propositional commonality. It cannot be a brute fact. The proposition *Smith punched Jones* is obviously a useful way of *identifying* that commonality. But that commonality *underlies* the fact that *Smith punched Jones* is true in both worlds. That fact cannot *constitute* the commonality. Otherwise, we strip the proposition of any anchoring in extra-propositional reality; and we deprive true propositions of the power to impose limits on how the underlying facts may be. So if *Smith punched Jones* is true in two different worlds, that indicates some more *fundamental* commonality. There is some more property P such that both worlds have P and such that it is a *consequence* of their having P that *Smith punched Jones* is true in each. The description "property had in common by all and only those worlds where *Smith punched Jones* is true" is a useful way of identifying that property. But the property is more fundamental than the corresponding proposition. And what enters into our analysis is the property itself, not the description of it. That is why our analysis is not circular.

§ One worry about our proposal is this. S doesn't have *Smith* as a member: instead, it has the property of being identical with Smith. By contrast, (Ap) has Smith, not the property of being identical with Smith, as a constituent. Therefore, S cannot be identical with (Ap).

There are (at least) two different ways we can respond to this. I don't think that either is more correct than the other.

For a long time, semanticists have insisted that Smith himself is a constituent of the proposition *Smith punches Jones*. But when we say such a thing, we are making an enormously theoretical claim. We don't have any direct understanding of what propositions are. When we make some claim about what the "constituents" of propositions are, it is to accommodate various data (about modality, analyticity, cognitive content, and the like). For example, we know that *that Smith punches Jones* couldn't possibly be true if there were no such thing as Smith. So this leads us to say that Smith himself must be a "constituent" of that proposition.

But if there should turn out to be some other way to accommodate this modal datum, then there would be no need to insist that Smith himself is a constituent of that proposition. We have identified (Ap) with S. S's being instantiated occurs exactly if (Ap) is true. So S's modal properties are in lock-step with (Ap)'s. Thus if we were to analyze (Ap) as being identical with S, that would be consistent with the relevant modal data. Since that analysis accommodates the relevant intuitions about modality, we (to that extent) have grounds for thinking it is the property of being identical with Smith, and not Smith *per se*, that is a constituent of (Ap). Our analysis does *not* accommodate the view that Smith himself is a constituent of (Ap). But *that* view is an entirely theoretical one; it is *not* a datum. It is surely not a *datum* that Smith is a constituent of that proposition. So our analysis cannot be thrown out on the grounds that it doesn't accommodate *that* view.

An analysis is meant to accommodate all the relevant *data* – not all the relevant theoretical superstructures that have grown up over that data. We can't say that Einstein's theory is wrong simply because it conflicts with the "datum" that Newton's theory is right. The latter is not a datum. We cannot say that our analysis of propositions is wrong simply because it makes the property of being identical with Smith, and not Smith himself, be a constituent of certain propositions. The idea that Smith himself is such a property is a bit of

theoretical superstructure; it doesn't have the same standing as any datum. And our theory is consistent with the relevant data.

§ There is another way to respond to the problem in question. We *could* say that our analysis *is* consistent with the idea that Smith is a "constituent" of (Ap). The idea would be this.

When we say that Smith is "constituent" of (Ap), that is really a metaphorical way of saying that the truth of (Ap) depends on some fact about Smith: some kind of relation of logical *dependence* holds between (Ap) and Smith. The sense in which Smith is a "constituent" of a proposition is entirely distinct from the sense in which something is a member of a set. The *membership* of S is not identical with the constituency of (Ap); and in order to do justice to the constituency of (Ap), we *must* assign members to S that are not constituents of (Ap). It is precisely *in order* to accommodate the fact that Smith is a constituent of (Ap) that we must make the property of being identical with Smith a constituent of S.

Let me clarify these obscure remarks. Suppose we take it as a given that Smith himself must be a constituent of (Ap). (I myself don't think that is a "given" at all. But let us leave that aside.) *By itself* that doesn't mean that our analysis is wrong. True – our analysis identifies (Ap) with some set S. True – S has as a member not Smith himself, but rather the property of being identical with Smith.

But it still doesn't follow that our analysis is inconsistent with the idea that Smith is a "constituent" of S. After all, we haven't yet analyzed the "constituency" relation yet. When we say that object O is a "constituent" of proposition P*, we are making a theoretical claim. We certainly cannot *see* that Smith is a constituent of some proposition the way we can *see* that a certain brick is a constituent of a certain house. When we talk about propositions, the term "constituent" is obviously being used in a metaphorical sense. Smith isn't a constituent of a proposition in the sense in which a leg is a constituent of table. Propositions are abstract objects. They don't have "constituents" in any straightforward sense. So even if we take it as a datum that Smith is a constituent of (Ap), we cannot say that our analysis is wrong *until* it is has been said what it is for something to be a constituent of a proposition. Once that has been said, *then* it can be determined whether our analysis is wrong.

But I believe that the concept of "constituency" can be delineated in a

way that reconciles our analysis with the aforementioned "datum." Suppose that (Ap) is true in some world w. According to our analysis, that means that the members of S are all *instantiated* in w. (Ap) is true exactly if all the properties in S are instantiated. One of those is the property of being identical with Smith. So, if our analysis is right, then (Ap)'s being true consists in various properties being instantiated, and Smith will be identical with one of the resulting *instances.*

So Smith is a "constituent" of (Ap) in the following sense. (Ap) is a set S of properties. For (Ap) to be true in w is for those properties to be instantiated in w. The property of being identical with Smith is one of those properties. So Smith is identical with one of those instances. Smith is a "constituent" of (Ap) in the sense that he is an *instance* of one of the properties that is a member of S.

In general, the constituency relation would amount to this. Let P* be an arbitrary proposition. P* is identical with some set S* of properties. O is a "constituent" of P* exactly if O is an instance of one of the properties that is a member of S*.

§ Let us take stock. Let (Ap) be the proposition *that Smith punches Jones.* According to our analysis, (Ap) is identical with a set of properties S. S has for its members various properties. One of the members is the property of being identical with Smith. Smith himself is not one of those members.

Given this, there are two ways we can deal with the widely held view that Smith is a "constituent" of (Ap). We can say:

(*) That view is wrong. It turns out that what is a member of S is not Smith himself, but the property of being identical with Smith.

The idea that Smith is a "constituent" of (Ap) is not a datum; it is a hypothesis – a piece of theoretical superstructure. One cannot deny that this table-leg is a constituent of this table: it is a datum; we see that it is true. But one cannot just see that Smith is a "constituent" of (Ap). The position that he is such a constituent is not a datum, but is rather a theoretical way of dealing with various data. Among these data are modal data, e.g. the datum that (Ap) couldn't possibly be true in a circumstance where there were no Smith. Our analysis of propositions accommodates that datum. And our analysis accommodates it by denying the theoretical (and obscure) view that Smith is a

constituent of (Ap). So, methodologically speaking, we have good reasons for denying the presumption that Smith is himself a constituent of (Ap).

The other position we can adopt is this:

(**) Smith is indeed a "constituent" of (Ap). Our analysis is consistent with that. Indeed, our analysis makes it clear what it is for Smith to be a constituent of (Ap). (Ap) is identical with a set of properties S. For (Ap) to be true is for all the members of S to be instantiated. One of those properties is that of being identical with Smith. In any world where (Ap) is true, Smith is an instance of one of the properties that is a member of S. Smith's being a "constituent" of P consists in its being the case that, in any world where (Ap) is true, Smith's is an instance of one of the properties that is a member of (Ap). In general, for O to be a constituent of a proposition P is for it to be the case that, in any world where P* is true, O is an instance of one of the properties that is a member of P*.*

I think that (*) and (**) are really the same view, the only differences being verbal.

§ Let us sum up. (Ap) is true exactly if all the members of S are instantiated. Further, (Ap) and S have the same decomposition. Therefore we might as well identify (Ap) with S and (Ap)'s being true with S's being such that its members are all instantiated.

At this point, there might seem to be a problem for our analysis. We just said that (Ap) and S have *exactly* the same decomposition: and surely they must, if S is to *be* (Ap). But S appears to have a constituent that Ap does not have. S comprises P; but (Ap) comprises only Smith, Jones, and the relation of punching – there is no P. So S and (Ap) have different constituents and are not identical.

I think this is false. I think that (Ap) *does* have a constituent corresponding to P. Smith, Jones, and the relation of punching are not the only constituents of (Ap). That proposition also has a *complex* constituent. Consider the verbal representation of that proposition: "that Smith punches Jones". The words "Smith punches Jones" seem to correspond to a single entity, albeit one that is, in some sense, composed of Smith, punching, and Jones. It is this

complex entity that gives (Ap) its truth-conditions. Smith, Jones, and the relation of punching contribute to (Ap)'s truth-conditions only by way of their involvement in this complex. So while it is true that one of S's constituents is something other than Smith, Jones, and punching, the same is true of (Ap).

§ One more application of our analysis might be of use. Consider the proposition

(#) *Plato snores.*

Let w be a world where (#) is true, and let w* be one where (#) is false. w is different from w*. The quantum in w is rippled differently: in w, there are events having a certain character and having certain origins; in w*, there are no such events. Let P be the property a world has iff its quantum is rippled in that way. (#) is true in a world exactly if P is instantiated in that world. This might induce one to say that (#) is *identical with* P and that for (#) to be true is identical with P's being instantiated.

But as it stands, this position is not quite acceptable. (#) has a unique, ultimate decomposition; and Plato and the property of snoring are uncontroversially among those ultimate constituents. (Whether they are *all* of them is another matter: I think not, for reasons I'll soon give.) Given this, suppose we identified (#) with P. In that case, if our identification is to be correct, then P must have the same decomposition as (#). But it is by no means clear whether this condition is met. If P is instantiated, let us say that there is a P-state of affairs. A P-state of affairs necessarily supervenes on other, lower-level states of affairs: states of affairs involving tissues in somebody's nasal passages, various interactions among air-molecules, respiratory and neurological processes, not to mention innumerable sub-atomic processes. Of course, the exact realization of any P-state of affairs may vary from world to world (or even within a world). But what is *not* thus invariant is that if there is a P-state of affairs, that is wholly in virtue of the existence of innumerable lower-level states of affairs meeting certain broad requirements. (It may not matter *exactly* how the air-molecules are distributed. But it matters that they go through *some* kind of passage in a body meeting *certain* requirements, with the result that noise of a *certain* kind is produced: within these limits, there is enormous latitude . But there are limits.) So a necessary condition for P's being instantiated is that various lower-level phenomena are instantiated. The

necessary conditions for a property's being instantiated presumably reflect facts about the structure of the property. Given this, it becomes at best an open question whether P decomposes neatly into Plato and the property of snoring. There is the distinct possibility that, if it decomposes at all, it decomposes into things much more basic than those two. Also, as we said before, there is no clear sense in which a property decomposes into others. By contrast, there is one clear sense in which *Plato snores* decomposes into Plato and the property of snoring. For these reasons, it would not be advisable to *identify* (#) with P.

Given this, here is what I would propose. (#) is identical with a set S *one* of whose members is P. The other members of S are Plato and the property of snoring. Remember, Plato is a property. Let Pp be this property, and let Ps be the property of snoring. If P is instantiated, then obviously Ps and Pp are also instantiated – any state of affairs sufficient for Plato's snoring is sufficient for the occurrence of Plato and for an occurrence of snoring. So *P is instantiated* entails *Pp and Ps are instantiated.* Thus, since the instancing of P is necessary and sufficient for the truth of (#), it follows that the instancing of P *and* Ps and Pp are also thus necessary and sufficient.

P is *true* exactly if all of S's members are instantiated. So P is identical with a set S whose members are Pp, Ps, and P; and P is true in w exactly if all of these properties are instantiated.

S has exactly the right decomposition. Among its ultimate constituents are Plato and the property of snoring. Further, if all the members of S are instantiated, that is necessary and sufficient for the truth of (#). So there is, so far as I can tell, no barrier to identifying S with (#) and to identity (#)'s being true with S's being such that its members are all instantiated.

§ Now we can say why, on our analysis, it is possible to distinguish between *analytically equivalent propositions.* As we noted, this is a distinction that many theories of the proposition have a hard time accommodating.

There is obviously a difference between the proposition

(#) *that Plato snores*

and the proposition

(##) *that Plato snores and triangles have three sides.*

One of these propositions is about triangles; the other is not. Our analysis easily accounts for the distinction. (#) is a set S of the kind just described. (##) is a set S* that has *some* members in common with S, but that also has other members. Among S*'s members are the property of being a triangle, the property of being a side, the property of being a trio (or of being thrice-instantiated). None of these things is a member of S. So and S and S* have different memberships, explaining why (#) and (##) have different decompositions.

There is more to say about the relation between (#) and (##). But it will have to wait until after we've dealt with molecular propositions. But given only what we just said, it should be clear why our analysis does, whereas the Lewisian analysis does not, account for the decompositional differences between (#) and (##). On Lewis' analysis, (#) is *identical* with some one property P such that P's being instantiated in w is necessary and sufficient for (#)'s being true in w. The trouble, of course, is that P's being instantiated in w is necessary and sufficient not only for (#) but also for (##) and for infinitely many other propositions. The difference between (#) and these other propositions is to be found, at least in part, in their decompositions; and our analysis has a place for those facts.

In light of these points, it should also be clear, at least to a point, why our analysis is able to distinguish between:

(i) *Alpha is a figure whose perimeter coincides with the class of all points equidistant from a given point in a plane*

and

(ii) *Alpha is a closed two-dimensional figure of uniform curvature.*

There is some property P such that if P's being instantiated in w is necessary and sufficient for (i) and (ii) *both* to be true in w. In fact, there is no conceivable ripple of w's quantum that could make the one true without also making the other true. So we cannot *identify* either of those propositions with the property of having a certain kind of quantum-ripple. But since the instancing of P is necessary and sufficient for the truth of each, it is not unreasonable to give it some place in our analysis of those propositions.

Given this, here is what I propose. (i) is identical with a set Si; (ii) is

identical with a set Sii. Both sets contain P. But the two sets don't have exactly the same constituency. Among the members of Si is the property of being a point; but this property is not a member of Sii. Among the members of Si is the property of having uniform curvature; this is not one of the members of Si. Each of (i) and (ii) is true exactly if all its members are simultaneously instantiated. (i) and (ii) are different propositions because, as just indicated, they have different constituencies. The truth of the one is necessary and sufficient for the truth of the other because the instancing of P is necessary and sufficient for the truth of both. The main point is that our set-theoretic analysis distinguishes between analytically equivalent but distinct propositions.

Another example might be in order. The proposition

(*) *a triangle is a triangle*

is not informative. But the following proposition is informative:

(**) *a triangle is a closed figure such that any two of its sides intersect, and such that not all three of them intersect.*

Here it might be appropriate to make a few general remarks. Again consider:

(#) *that Plato snores*

There is reasonable agreement that Plato and the property of snoring are constituents of (#). Our analysis is consistent with that. But our analysis might *seem* inconsistent with another obvious fact about (#). The two constituents just mentioned *combine* to form some larger constituent – one corresponding to the words "Plato snores" in the verbal expression of (#). Here is the trouble. We identified (#) with a set S whose members are Ps, Pp, and P. On our analysis, P is an *ultimate* constituent of S – it is no less a basic constituent of S than are Ps and Pp. When we look at (#) (or, rather, at its verbal representation), it looks as though the only *basic* constituents of that proposition are Plato and the property of snoring. To be sure, (#) does have a third constituent – one that corresponds to some kind of putting together of the words "Plato" and "snores". But, by all appearances, that *third* constituent is

built out of the other two: it mustn't be *basic.* But our analysis makes it basic: P is as ultimate and indivisible a constituent of S as either of Ps or Pp. Therefore, one might conclude, our analysis is wrong.

I think that this objection to our analysis involves a failure to distinguish propositions from their verbal representations. It also involves a failure to appreciate how much we rely on sheer metaphor when we discuss propositions. Finally, that objection involves a failure to distinguish the psychological facts relating to the grasping of a proposition from the facts relating to the structure of the proposition itself.

There is no doubt that (#) is in some sense "put together out of" Plato and the property of snoring. But there is also no doubt that this putting together has not the slightest resemblance to the putting together of a car or a building. The property of snoring is an abstract object; it cannot be put next to, or on top, or underneath anything. And given any two spatiotemporal entities, no matter how you "put them together" – no matter what kind of glue you use, no matter what kind of forklift you operate – you will not make a proposition out of them. So as for the idea that the third constituent of (#) – the "complex one" – has to result from some kind of juxtaposition of the other two – that idea is a complete non-starter. That third component will not divide into Plato and the property of snoring in anything like the way that table divides into legs and a plank.

I believe that our analysis actually makes it clear what the *actual* sense is in which Plato and the property of snoring combine to form a third constituent. If P is instantiated, that is *ipso facto* sufficient for the instantiating of Pp (i.e. Plato) and of Ps (i.e. the property of snoring). There is obviously a dependency relation among these various properties. These dependencies are given by entailments. Now properties cannot literally be parts of other properties. A property is an abstract object, and cannot be a part of anything in the sense in which a leg is a part of a body. When we talk about properties being "parts" of other properties, we must surely be talking about relations of *dependence*. These relations are often, though not always, embodied in our verbal definitions: a "triangle" is a "three-sided closed, straight-edged, planar figure". We might say that the property of being a side is a "part of" the property of being a triangle. But we are using the word "part of" metaphorically here, since abstract objects, like properties, don't have parts in any ordinary sense of the word. The *non*-metaphorical meaning here must be this: *if* x has the property of being a triangle, *then* x has the property of having

straight-sides. So when we talk about properties being parts of other properties, we must, I think, be talking about relations of dependence or entailment holding among instances of those properties.

That relation does seem to hold among P, Ps, and Pp. If P is instantiated, so *ipso facto* are the other two properties – just as when the property *triangle* is instantiated, so are *side*, *three*, *straight* and so on.

Of course, in the verbal representation of (#), the part corresponding to P is literally composed out of the parts corresponding to Ps and Pp. But that obviously doesn't mean that P itself is literally composed out of those properties. Further, that fact about language actually supports our analysis, if only weakly. The kind of dependence-relation just discussed is often verbally expressed through the sort of *verbal* juxtaposition in question: the expression "three-sided closed, planar, straight-edged figure" picks out the property of triangularity. This might make it seem as though the property picked out by that long expression literally *composed* of the properties *straight-edged*, *planar*, and so on: they are literally *parts* of the property of triangularity. But the only sense in which the properties *planar, straight-edged*, and so on, are *parts* of the property of triangularity is that some *dependence* relation holds among these properties: *if* a thing is a triangle, that entails that it is also planar and straight-edged, and so on. So the *verbal* decomposition of "Plato snores" into "Plato" and "snores" corresponds not to a relation of *part*-hood, in any literally sense, but to a relation of dependence. And that is precisely the relation that holds between P and Ps and Pp.

We are inclined to say that Ps and Pp are the only *basic* components of (#), and that any other components are "built out of" those two. The only sense that can be attached to this is: if (#) comprises any third component, then there is some kind of asymmetrical relation of dependence—necessitation or entailment -- between that third component and the other two. On *some* interpretation of the expressions "less basic than" and "composed out of", the property picked out by "three-sided closed, planar, straight-edged figure" is less basic than, and is composed out of, the properties picked out by "three-sided", "closed", and so on. But all this means is that *x is triangle* entails *x is straight-edged (closed)* but not vice versa. If we insist on understanding the relation of "composition" in more literal terms, we will inevitably fail. Obviously the aforementioned relation of dependence *does* hold between P, on the one hand, and Ps and Pp on the other. If P is instantiated, that requires that Ps (Pp) be instantiated, but not vice versa. Our analysis of (#) is consistent

with that: on our analysis, the component whose verbal expression consists of *both* "Plato" and "snores" stands in exactly that relation of dependence with respect t the component whose verbal expression is "Plato" and also with respect to the component whose verbal form is "snores". So our analysis *does* capture the only meaningful sense in which certain components of that proposition are more "complex", or our "built out of", others. It is true that, on our analysis, P is an ultimate constituent of (#). But that is in no way inconsistent with standing in the dependence -relations just described with respect to Pp and Ps; and, as we've seen, to the extent that it has meaning to say that P is a "derived" constituent of (#), and not a basic one, what is meant can only be that P stands in the afore mentioned dependency relations with respect to these other constituents. Since our analysis captures that fact, our analysis captures the only meaningful sense in which certain constituents of that proposition are "built out of", and are thus "less basic than", others.

§ We must take care to distinguish facts relating to the *epistemology* of propositions from facts relating to the structure of propositions themselves. A failure to make this distinction is, I believe, abetted by a tendency to put too much stock in the verbal representations of propositions. When you read the *sentence:*

(#s) "Plato snores",

you "piece together" the meaning of the whole sentence on the basis of the meanings of its parts. This, of course, might lead one to think that the meaning of the whole is put together out of the meanings of the parts in something like the way that a table is put together out of legs and a slap of wood.

When we say that you "piece together" the meaning of the whole on the basis of the meanings of the parts, the only thing this could mean is that you *figure out* the meaning of the whole sentence on the basis of the meanings of the parts. When we say that you figure this out, we can only be saying that you have hit on some dependency relation holding between the meanings of the parts and that of the whole. Once again consider the expression: "three-sided closed, planar, straight-edged figure". That expression, as a whole, picks out the property of triangularity. On the basis of the meanings of the parts of that expression, plus a small amount of mathematical acumen, you can figure

out what the expression as a whole picks out. But this obviously doesn't mean that the referent of the whole expression is, in any literally sense, put together out of the meanings of the parts. You can *derive* the meaning of the whole on the basis of the meanings of the parts; and the fact that this can be done embodies the fact that there is a dependency relation between the thing picked out the by whole and the things picked out by the parts. But that is the *only* sense in which the thing picked out by the whole is "composed of" the things picked out by the parts. The *only* sense in which it is meaningful to say that one property is "composed out" of another is to say that there is some kind of dependency relation of the kind described.

Let us sum up. The fact that we *figure* out the meaning of a whole expression on the basis of the meanings of the parts means only that some (asymmetrical) dependency relation holds between the meaning of the whole and the meanings of the parts. It does *not* mean that the whole is, in any other sense, less basic than the parts. Indeed, in this context, the only meaningful sense that can be attached to phrases like "less basic than" and "is a complex consisting of" concerns relations of dependency. In the only sense of "more basic than" that has any significance in this context, our analysis is consistent with the fact that some constituents of propositions are "more basic than" others. In the only relevant sense of "more basic than", our analysis captures the fact that Plato (Pp) and the property of snoring (Ps) are more basic than P. Our analysis captures the fact that P's being instantiated requires Pp's (Ps's) being instantiated, but not vice versa. It is true that, on our analysis, P and Ps and Pp are all equally members of S: from that viewpoint, none is more basic than any of the others; from that viewpoint, there is no stratification. S is a kind of democracy, and P is as lowly as Ps and Pp. But that doesn't matter. The only relevant sense in which P is to be put on a higher level than Ps or Pp is that that instantiating of P instantiated necessitates that of Ps (Pp) but not vice versa. Our analysis captures that fact in spades. Every other kind of stratification or sub-ordination is irrelevant metaphor. So whatever our feelings about it might be, there is nothing logically wrong with identifying (#) with a set in which P, Pp, and Ps are all on an equal footing: their *inequality* – the asymmetry of the entailment relations holding among them -- is inseparable from them, and needn't be buttressed by encasing them in some stratified structure.

§ These points help us deal with another important fact. Propositions

are *structures;* they are *orderings*, not heaps, of constituents. But if we focus too much on language , we are likely to misunderstand the nature of this ordering.

Consider the sentence "Smith punches Jones". The order of the words has nothing to do with the order of the corresponding entities in the proposition. In some languages, the order of the nouns is reversed; in others, there are (virtually) no constraints on word-order. The sense in which Smith, Jones, and the relation of punching are "ordered" in

(Ap) *Smith punches Jones*

has nothing to do with order in the spatio-temporal sense.

Once again, the kind of order in question has to do with relations of dependence – relations of entailment. It is true that, in (Ap), Smith has one "position" and Jones has a different one. In

(A*p) *Jones punches Smith*

Smith and Jones have switched positions. What, in non-metaphorical terms, what does this all amount to? We saw earlier that, when we are discussing propositions, facts about "composition" and "stratification" are really facts about dependence. The same thing is true here: facts about the "ordinal" properties of propositions are really facts about entailment relations. (Ap) entails that Smith punches something; (A*p) does not entail this. (Ap) entails that Jones is punched by something; (A*p) does not entail this. And so on. When we say that, in (Ap), Smith is in one "position" and Jones is another, what we are saying is that, if (Ap) is to be true, then Jones must be doing the punching and Smith must be receiving it. We are talking about what is entailed by the truth of the proposition. So we are talking, not about position in any literal sense, but about relations of dependence and entailment. The ordinal properties of propositions are fixed by their inferential properties.

Consider (Ap). We pre-theoretically know that Smith is in the "subject" position and Jones is in the "object" position. If we identify (Ap) with something x, then we must be able to retrieve the ordinal information that we pre-theoretically have about (Ap) from x. The ordinal properties of (Ap) must be fully "coded" in x. In general, an analysis of what propositions are must fully account for their ordinal properties.

I believe our analysis satisfies this requirement. According to us, (Ap) is identical with a set S whose members are the following properties (remember that individuals are in fact properties)

Ps: Smith
Pj: Jones
Pp: the relation of punching
P: the property that a world has if, in it, Smith is punching Jones.

And for S to be true is for all of those properties to be instantiated. Of course, if all of those properties are instantiated, then P is instantiated. If P is instantiated, there are various consequences: Smith is punching something, Jones is being punched by something, and both punchings are one and the same. We saw a moment ago that the ordinal properties of (Ap) lie in these consequences. So by virtue of identifying (Ap) with S, and by identifying the truth of (Ap) with the instantiating of all of S's members, we have ensured that S codes all the ordinal information characteristic of (Ap).

§ Here it might be appropriate to point out what are, I believe, some virtues of our analysis. Remember that theory (i), and also Wittgenstein's "picture theory, couldn't account for the existence of false propositions, since they identify propositions with their truth-makers. On our account, the *truth* of a proposition requires the existence of some spatio-temporal state of affairs. But the proposition *itself* requires only the existence of properties: it requires only the existence of *ways* that mass-energy might be displaced.

Molecular Propositions

Let us discuss *molecular* propositions: these are propositions that are formed out of other propositions, e.g. *Bob is tall and Sam is short*. In keeping with tradition, I will describe quantified propositions (e.g. *something is short*) as "molecular".

A point about methodology is in order. When dealing with atomic propositions, we defined truth in terms of instantiatedness: a proposition is a set of properties, and that proposition is true iff all its members are instantiated. We must treat molecular propositions in the same way. We cannot define *proposition* and *truth* one way for atomic propositions and another way

for molecular propositions. That would be absurd. Propositions are propositions. Truth is truth. But this need for uniformity requires that, initially, we say a few counterintuitive things.

Consider the proposition:

(jsw) *That John believes that snow is white.*

This is a proposition about a proposition; it attributes a property – being believed by John – to *that snow is white.* Obviously that property and that proposition are constituents of jsw. So jsw is identical with a set S among whose members are that property and that proposition.

Here a refinement is in order. We have identified propositions with sets of properties: it is crucial to our analysis that every component of a proposition be something that can be instantiated. But, as we've analyzed it, the proposition *that snow is white* is not exactly a property; it is a set. (I don't think that properties *are* sets.) In order to provide a uniform analysis, I propose this. What is a constituent of jsw is the property of being identical with *that snow is white.* That property is instantiated exactly if that proposition *exists.* We must be careful here: the property of being identical with *that snow is white* can be instantiated without that proposition's being *true.* For the instancing of that property, it is necessary and sufficient only that *that snow is white* exist. To sum up, jsw is a set S; and among the members S are the property of being believed by John and also the property of being identical with *that snow is white.*

But those cannot be the only members of S. Both of those properties can be instantiated *without* John's believing that snow is white. The property of being identical with *that snow is white* is automatically instantiated. If John believes that grass is green (or even that snow *isn't* white), then the property of John's believing something is instantiated. So the properties in question can be instantiated with jsw being true. S must comprise some other constituent. This other constituent is not hard to identify.

In worlds where John believes that snow is white, the quantum is rippled one way; in other worlds, the quantum is rippled differently. Let P be the property that a world has if, in it, the quantum is rippled in the way just mentioned. So P is instantiated in w exactly if, in w, John believes that snow is white. Among the members of S are P. Of course, if P is instantiated, then so are the properties that are members of S. Obviously if P is instantiated, then

John believes something or other (so the property of being believed by John is instantiated). And the property of being identical with the proposition *that snow is white* is instantiated under any circumstances. Thus if P is instantiated, so are the other members of S. And jsw is true exactly if P is instantiated. So jsw is true exactly if all three members of S are instantiated. Thus we identify jsw with S, and we identify S's being true with all its members being instantiated.

The question arises: Why not just say that P is the *sole* member of S, since P's being instantiated is necessary and sufficient for the truth of S? We've already seen the answer to this. It is pretty clear that jsw comprises a distinct constituent corresponding to the property of being believed by John, and it also comprises a distinct constituent corresponding to the proposition *that snow is white* (and, thus, to the property of being identical with that proposition).If we identify jsw with a set, then we need the membership of that set to clearly reflect these facts about jsw's constituency. If that set's sole member is P, then those facts are eclipsed; the set doesn't have the right decomposition.

A given propositions is "composed" of various things. We saw earlier that this composition-relation is best understood in terms of relations of dependency or entailment holding among those constituents. It is pretty clear that jsw comprises one constituent corresponding to the words "john believes", another constituent corresponding to the words "that snow is white". It is *also* clear that jsw comprises some more "complex" constituent corresponding to the juxtaposition of "John believes" and "that snow is white". If that more complex constituent exists, then so does each of the simpler ones; but both of the simpler ones can exist without the complex one existing. The thing corresponding to "John believes" can exist simultaneously with the thing corresponding to "that snow is white". But that doesn't guarantee the existence of the thing corresponding to the juxtaposition of those expressions. So an asymmetrical relation of dependency holds between that more complex constituent and those less complex ones.

Exactly these points apply to the members of S. If P is instantiated, then so is each of the two properties just mentioned. But both of those other properties can be instantiated with P being instantiated. So there is a dependency relation between P and each of the other two properties just mentioned; and that relation perfectly parallels the one discussed a moment ago – both relations are asymmetrical in precisely the same way.

Given this, it becomes clear why we must identify jsw with S. S comprises three constituents: one corresponding to the proposition *that snow is white*; another to the property of being believed by John; and a third corresponding to something that, in the relevant sense, combines the two. (Remember that the kind of combining in question consists in the holding of dependency relations: and the right kinds of dependency relation holds among the three things just discussed.) So S has the same constituency as jsw. Further, jsw is true exactly if all three members of S are instantiated. So we can identify jsw's being true with its being the case that all the members of S are instantiated – roughly, we identify truth with instantiatedness. And we identify jsw with S. In this way, we do justice to the decompositional fine-grain of jsw, and also to its truth-conditions.

At this point, I must address what many will find to be a worrisome feature of our analysis. We have identified propositions with sets. Thus the proposition:

(jsw) *John believes that snow is white.*

is a set. We decided that what is a member of that set is not, strictly speaking, the proposition *that snow is white*, but rather the *property* of being *identical with that proposition.* Obviously this proposal is counter-intuitive. One would think that the proposition itself ought to be a member of that set.

There are a couple of things to say here. First, on one delineation of the expression "constituent", we are making the proposition *that snow is white* be a constituent of (jsw). But, even if we leave that aside, I think that the just mentioned counter-intuitiveness of our analysis is simply another embodiment of some very crude and false theories that we tend to have about the sense in which things are "constituents" of propositions. *That Plato snores* is not built out of Plato the way a building is made out of bricks. (jsw) is not built out of *that snow is white* the way my car is built out of metal parts. As we saw earlier, whenever we relate the notions of "constituency", "composition", and "part" to propositions, we are really talking about relations of logical dependence and entailment. On our analysis, the truth of (jsw) absolutely *depends* on some fact about the proposition *that snow is white.* On our analysis, (jsw) is true exactly if *that snow is white* has the property of being believed by John. In so far as our analysis of (jsw) is consistent with that fact, it fully accommodates the fact that *that snow is white* is "constituent" of (jsw)

– for that "constituency" relation just *is* the dependency relation discussed a moment ago. (In any case, that is one way to delineate the concept of "constituency", as it relates to propositions.) Now *in order* to accommodate that fact, we had to identify jsw with a set S that comprises, not *that snow is white*, but instead the property of being identical with that proposition. But this is in no way inconsistent with the truth expressed by statements like "the proposition that snow is white is a constituent of jsw": for on any interpretation in which such a statement is meaningful, let alone true, that statement amounts to an affirmation of the dependence relation described a moment ago; and, as we just saw, our analysis fully accommodates the existence of that dependence relation. So by identifying jsw with a set S one of whose constituents is the property of being identical with *that now is white*, our analysis actually *captures* the truth of the statement that *that snow is white* is a constituent of jsw, in so far as that statement is meaningful and true.

What we said about jsw applies, with only a few obvious changes, to all molecular propositions. Consider:

(sn) Sam is not short.

This attributes falsity to the proposition *that Sam is short.* So among sn's constituents are that property and that proposition. For reasons discussed earlier, in connection with jsw, it would be preferable to say that the property of being identical with *that Sam is short* is a constituent of that proposition. So sn is a set S that has as members (inter alia) the property of being false and the property of being identical with the proposition *that snow is white.* (The property of falsity is instantiated in a world w iff some proposition is false in w. Presumably that property is instantiated in every world. But that needn't concern us here.)

Like any other proposition, sn is true iff all the members of S are instantiated. But obviously sn could be false even if both the property of being false and the property of being identical with *that Sam is short* were instantiated. (After all, those properties are instantiated in w if it is false in w that Smith is bald and true that Sam is short.) So S must comprise some third constituent.

In light of our earlier points, this third constituent is not hard to identify. Consider the way the quantum is rippled in a world in which Sam is not short, i.e. in which Sam is of average or above average height. Let P be the

property that all and only such worlds have. P's being instantiated in w is necessary and sufficient for sn's being true in w. So P must be a member of S. But we don't want to make P be the only member of S. For now familiar reasons, making P be the sole member of S wouldn't do justice to the decomposition of sn. But we are being true to that decomposition if we make P *and* the other two properties (that of being false and that of being identical with *that Sam is short*) be the membership of S. So has the same decomposition as sn. Further, sn is true exactly if all of those three properties are instantiated. So we identify S with sn, and identity sn's being true with S's being such that all its members are instantiated.

§ Before dealing with quantified propositions, let us deal with conjunctions. Consider:

(sb) *That Sam is tall and Bob is short.*

First of all, this proposition is true exactly if the two constituent propositions are true. So the constituent whose verbal representation is "and" can be treated as a property that ascribes truth to a pair of propositions exactly if both members of that pair are true. This analysis is, I will argue later, equivalent to the familiar idea that "and" denotes a function that assigns truth to a pair of propositions exactly if both members of that pair are true.) Let AND be that property. The property AND is instantiated in a world w exactly if some pair of proposition is true in w. (Presumably AND is instantiated in every world. But this needn't concern us.)

Obviously (sb) has AND as a constituent. For now familiar reasons, (sb) also comprises the property of being identical with *that Sam is tall* and also the property of being identical with *that Bob is short.* Let Pst and Pbs be these two properties. So, for now familiar reasons, sb is identical with a set S whose membership includes AND and Pst and Pbs. But – once again for familiar reasons – those cannot be the only members of S. After all, all of those properties could be instantiated with sb's being true. So S must contain a fourth constituent. Consider the way that quantum is rippled in a world where it is the case both that Sam is tall and Bob is short. Let P be the property had by all and only such worlds. sb is true in a world w exactly if P is instantiated in w. (Of course, is P is instantiated in w, then so are AND and Pst and Pbs.) If we suppose that S comprises P, then sb is true exactly if every member of S is

instantiated. Further, S has the same decomposition as sb – both comprise distinct and separable parts corresponding to the property picked out (in English) by "and", and also to the proposition *that Sam is tall* and the proposition *that Bob is short.* So we might as well identify sb with S, and identify the truth of sb with its being the case that every member of S is instantiated.

Similar discussions provide the analysis of propositions like *Bob is short or Sam is tall, Bob is short because Sam is tall, Bob is short if Sam is tall,* and so on. There are no conceptual obstacles here, though some care must be taken in choosing the constituents corresponding to the connectives.

§ *Quantified Propositions*

In keeping with tradition, I regard quantified propositions as being molecular. The proposition *nothing snores* is in the same category as *Bob snored and Fred snores,* and not in the same category as *Bob snores.* I adopt this view partly because, as we will see, it allows for a smooth and uniform semantics. (So the reasons are somewhat like the reasons for seeing n^0 as being equal to 1, for any natural numbers n). But I also think that, at bottom, quantified propositions really are in the same category as conjunctions, disjunctions, and so forth. They are meta-propositions: propositions about classes or sequences of propositions. (I hope that reasons for this become clear.)

Let us begin our treatment of quantified proposition with:

(ns) *that nothing snores.*

I think it is important to begin with negative existentials, since they make obvious certain features of quantified propositions hidden by existentials and universals. Following Frege, I see (ns) as saying that every proposition having the form *x snores* is false. For now, I would like permission to treat the form *x snores* (sometimes referred to as a "propositional function" or "open proposition") as an entity in its own right. In a moment, we will discuss what kind of entity it is.

On the face of it, (ns) has (inter alia) two constituents: the propositional form just mentioned, and the property of being universally false, i.e. the property had by a thing x if every instance of x is false. So the form *x is*

a square circle has that property, since every instance of that form is false. Let us refer to that property as UF. UF is instantiated in a world w iff there is some propositional form such that, in w, every instance of that form is false. (Presumably UF is instantiated in every possible world. But, once again, this needn't concern us.) For now familiar reasons, we must identify ns with a set S whose membership includes (inter alia) UF and the form *x snores.*

There is a caveat. We have said that a proposition is a set of propositions, and is true exactly if all those properties are instantiated. Here is the problem. Let NS be the set with which (ns) is identical. We don't want to say that the property of snoring is a member of that set. Why not? Because if that property is instantiated, then (ns) is automatically false. So we don't want to say (ns) has that property as a member.

It is not hard to deal with this. What is a member of S is not exactly the form *x snores*, but rather the property of being *identical* with that form.

This procedure is exactly analogous to one we used earlier. We decided that

(stbs) *that Sam is tall and Bob is short*

is identical with some set S; and decided that what is a member of S is not exactly the *that Sam is tall*, but rather the *property* of being identical with *that Sam is short.* (Let Pss be that property.)

We saw that, contrary to first appearances, this procedure actually *accommodates* the fact that *that Sam is tall* is a "constituent" of (stbs), at least on one viable delineation of the term "constituent". For that procedure *accommodate* the dependence relation that holds between (stbs) and *that Sam is tall*; and, as we saw, that dependence-relation is really identical with the constituency-relation in question. Further our decision to make Pss be a constituent of S enabled us to extend our analysis of truth as instantiatedness to molecular propositions. That decision enables us to capture both the decomposition and the truth-conditions of (stbs): and those are the two important desiderata.

Let us return to

(ns) *nothing snores.*

We want our analysis of (ns) to do justice both to its decomposition and to its truth-conditions. This can be done, provided we apply the procedure (mutatis mutandis) just described. We identify ns with a set S that comprises (inter alia) UF and also the property of being identical with the form *x snores.*

Before we continue, a somewhat delicate point is in order. As we will

see, propositional forms probably *are* properties. I believe that *x snores* is the property had in common by all and only *that John snores, that Mary snores, that Frank snores*, and so on. I believe that an *instance* of that form is some *specific* proposition of that form. So an *instance* of *x snores* is a constituent of *that John snores.*

But *that John snores* is not an *instance* of the property of being identical with *x snores*. An instance of the *latter* property is the form *x snores* itself.

An example may clarify my meaning. An instance of the property *even number* is the number four. But the number four is not an instance of the property *identical with the property of being an even number.* After all, the number four is not *identical* with the property of being an even number. Similarly, *that John snores* is not an instance of the property *identical with the propositional form: x snores*, even though it is identical with an instance of the form *x snores.*

To sum what we've said thus far: ns is identical with a set whose membership includes the property of being identical with the form *x snores* (let Pxs be that property) and also includes UF. I propose that, by doing this, we can account for the truth-conditions and the decomposition of ns.

But, again for familiar reasons, S cannot comprise *only* UF and Pxs. After all, *those* two properties could be instantiated in a world w where ns was false. Suppose that, in w, the proposition *nothing plays tennis* is true. In that case, UF is instantiated in w. Further, suppose that Mary snores in w. In that case, Pxs is instantiated in w. (Of course, Pxs can be instantiated in a world even if *nothing* in that world snores. For, it seems to me, even in a world where nothing snores, the property of being *identical* with the form *x snores* exists. But, once again, we needn't dwell on this.) So both UF and Pxs can be instantiated in a world without ns being true. So ns cannot be identical with a set that contains *only* UF and Pxs.

So given that we are identifying ns with a set, that set must contain some third constituent. Again, this is not hard to identify. Consider the way the quantum is rippled in a world where something snores. Let P* be the property that all and only such worlds have. Now let P be the property that all and only worlds lacking P* have. So a world has P iff its quantum is appropriately *un*rippled. ns is thus identical with a set S that comprises UF and Pxs *and* P. ns is true exactly if P is instantiated. Of course, if P is instantiated, so *a fortiori* are UF and Pxs. It follows that ns is true exactly if P and Pxs and UF are

instantiated. But, for now familiar reasons, we don't want to say that S's sole member is P: that wouldn't do justice to the compositional structure of S. So we identify ns with a set S containing all three of the properties just mentioned; and we identify ns's being true with S's being such that all its members are instantiated.

§ Our analysis made heavy use of the idea that there is such a thing as the propositional *form* x snores? Now we must answer the question: what is that thing?

Consider the class of propositions containing all and only *that John snores, that Mary snores*, and so on. The form *x snores* is the property had in common by all and only those propositions. In light of what we said earlier, we can give a precise statement of what this property is. Remember that the proposition *John snores* is identical with a set S whose members are: the property of being identical with John; the property of snoring; and the property P had by world w exactly if, in that world, John snores. An exactly similar analysis applies to *that Mary snores* and *that Ethel snores* and so on. So the property of being identical with the form *x snores* is one and the same with the property of being a set S such that, for some individual x, S contains (i) the property of being identical with x; (ii) the property of snoring; and (iii) the property P had a by world w exactly if, in that world, x snores.

§ The proposition:

(es) Everything snores

is dealt with in a way exactly analogous to the way we dealt with (ns). First of all, (es) is true if the propositional form *x snores* is "universally true", i.e. is such that each instance of that form is true. Let UT be the property of universal truth. We identify (es) with a set S that contains UT and, for reasons already seen, that also contains Pxs (the property of being identical with the property *x snores*). For now familiar reasons, S must contain a third constituent. Consider the way the quantum is rippled in a world w in which everything snores. Let P be that property. S contains P and UT and Pxs. (es) is true exactly if P is instantiated. If P is instantiated, then so are Pxs and UT. So (es) is true exactly if all of P and Pxs and UT are instantiated. Thus (es) is true exactly if all the members of S are true. Thus we identify (es) with S, and

identify the truth of (es) with its being the case that all the members of S are instantiated.

§ Let us now consider the proposition:

(ss) *something snores.*

Here it is tempting to say the following:

It is obvious what that proposition is. It is identical with the property of snoring. And that proposition is true exactly if that property is instantiated. So (ss) fits perfectly into our view that propositions as sets of properties and that for a proposition to be true is for the corresponding properties to be instantiated.

But such a view is not correct. Presumably *something snores* has the same basic structure as *nothing snores* and *everything snores.* For reasons we have seen, the latter have as members, not the property of snoring, but rather the property of being identical with that property. We want (ss) to have a structure analogous that of those other two propositions.

Also, (ss) has a certain amount of decompositional structure. The property of snoring doesn't have any decompositional structure, at least not in the sense in which (ss) does.

Given these points, this is what we say. (ss) is identical with a set (SS). (SS) has the property of being identical with the property of snoring as a member. (SS) also has the property of being instantiated as a member. Let (PT) be this property.

(ss) also has a third member. Consider the way the quantum is rippled in a world where something snores. Now consider the property common to all and only worlds having that property. Let (s*) be that property. For exactly analogues of reasons considered earlier, in connection with (ns) and (es), (SS) has (s*) as a member.

So the proposition *something snores* is identical with a set that has the following three things as members: the property of being identical with the property of snoring; (s*); and (PT). That proposition is true in a world exactly if each of those properties is instantiated in that world.

This analysis is, admittedly, counter-intuitive. But it is demonstrably consistent with both the truth-conditions had by *something snores* and also

with facts about its decomposition.

Here one consequence of our analysis should be made explicit. According to our analysis, the proposition *that John snores* is identical with a set that has the property of snoring as a member. (And our analysis also requires that that property be instantiated if that proposition is correct.) But on our analysis, (ss) is identical with a set that does *not* have that property as a member. And this will strike many as odd, and even unacceptable.

Let us recall a point made earlier. The *John snores and Fred snores* is *obviously* built out of other propositions. It is *obviously* "molecular". Traditionally, semanticists have regarded quantified propositions (like *something snores*) as also being molecular, even though they are not, strictly speaking, built out of other propositions. I think that this tradition embodies a real insight into the nature of quantified generalization. When you say "something snores", you are, arguably, making a statement about a *class* of propositions: the class of propositions of the form *x snores.* You are saying that at least one such proposition is true. So quantifications are meta-propositions. For reasons given earlier, the same is true of conjunctions, conditionals, negations, and the like. The proposition *if grass contains chlorophyll, then grass is green* affirms the existence of a relation between propositions. It is thus a meta-propositional proposition. The right definition of "molecular proposition", then, is probably not "proposition built out of other propositions", but rather "proposition that concerns propositions". As we just saw, quantified generalizations fall into this category, as do our paradigms of molecular propositions (conjunctions, conditionals, disjunctions, and so on). So the tradition of assimilating those two classes into a single proves correct or at least theoretically very convenient.

§ Now we can give fairly precise answers to two questions. First, what is it for E to "denote" O? Second, what is the semantic function of *grammar*?

There are two delineations of the term "denote". One of these, E denotes O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition P such that P's being true involves the existence of a state of affairs such that O is a constituent of that state of affairs. This is the sense in which

"Smith" denotes Smith. Consider the sentence "Smith punches Jones." Suppose that (Ap) – the proposition encoded in a token of that sentence – is true. P is true in virtue of the fact that there is a state of affairs

that (inter alia) has Smith as a component.

Let us refer to this as "denoting1". What does the expression "red" denote1? The obvious, but wrong, answer is: "the property of redness". The right answer is this. A token of "red" denotes1 an *instance* of that property. Consider a token of the sentence "that car is red". If true, that token encodes a proposition P such that P holds in virtue of the existence of a state of affairs of which some constituent is an instance of redness.

To understand the other delineation of "denote", we must remember that each proposition is identical with some *set.* On this delineation, E denotes O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition P such that the set with which P is identical has O as a member. Let us refer to this relation as "denoting2". Consider a toke of "Smith is tall". The occurrence of "tall" denotes2 the *property* of tallness, not some instance thereof. And the occurrence of "Smith" denotes2 the property of being identical with Smith, not Smith himself (i.e. not some instance thereof).

Here is the general rule.

E denotes1 O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition P such that, if P is true, it is true in virtue of the existence of a state of affairs one of whose constituents is O.

E denotes2 O exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition P such that O is one of the members of the set with which P is identical.

§ What do grammatical inflections do? "Smith punches Jones" encodes a proposition; "Smith, the relation of punching, Jones" does not. The first has the right grammatical inflections; the second does not. The difference between a sentence-token and an unmeaning heap apparently lies in facts about grammar. Grammatical facts code the relevant ordinal information.

We've seen what that information is.

(Ap) The proposition *Smith punches Jones.*

Ps: the property of being identical with Smith.
Pj: the property of being identical with Jones.
Psx: the property of being a thing x such that Smith punches x.
Pxj: property of being a thing x such that x punches Jones.
Pp: the relation of punching.

P: as previously defined.

(Ap) is a set S whose members are exactly the properties just listed, and for (Ap) to be *true* is simply for all of its members to be instantiated.

Let t* be a token of "Jones punches Smith", and let (Bp) be the proposition encoded therein. (Bp) and (Ap) will have different constituents. t and t* differ in the order in the component expressions are arranged. This is a grammatical difference; it is the way in which, in English, facts about *case* are expressed. In other languages – Latin, Russian, Greek – those facts are coded primarily in word-endings, not (so much) in facts about word-order. In English, word-order does exactly what is done, in languages like Russian and Latin, by patently grammatical inflections. So word-order is a grammatical device *par excellence.*

What does word-order (in English) do? The set with (Bp) is identical will have different members from the set with which (Ap) is identical. For example, (Ap) has as a member:

Psx: the property of being a thing x such that Smith punches x. property of being a thing x such that Smith punches x.

(Bp) does not have Psx as a member. (Bp) has as a member:

Pjx: the property of being at thing x such that Jones punches x.

(Ap) does not have Pjx as a member:

Grammatical inflections, then, do have an extremely significant semantic role: they determine what will be the constituents (on any delineation of that term) of the propositions meant by sentence-tokens. In t, the facts about word-order tell you (inter alia) that Psx is a member of the corresponding proposition. In t*, facts about word-order tell you (inter alia) that Pjx is a member of the corresponding proposition.

Are grammatical inflections referring terms? There are two ways we can go about answering this question.

Grammatical inflections do not *by themselves* refer to anything; but they *assign* referents to sequences of expressions on the basis of the referents

of those expressions. Consider t again. As a whole, t denotes the proposition (Ap). The grammar of t – the word-order, the inflection on the verb – do not themselves denote anything. But, on the basis of the referents of "Smith", "love", and so on, they assign referents to various complex expressions. For example, the proposition (Ap) is assigned to the whole sequence. The property Psx is assigned to the complex expression "Smith punches", on the basis of the referents of "Smith" and the root-verb "punch". The property Pxj is assigned to "punches Jones" on the basis of the referents of "Jones" and the root-verb "punch".

Very roughly, the general idea would be this. Grammatical inflections do not themselves refer. But when a grammatical inflection occurs in a sequence of referring terms, that occurrence assigns a referent to that sequence as a whole on the basis of the referents of the expressions composing that sequence.

There is another way to look at facts about grammar. The idea would be the grammatical inflections *do* refer; but what they refer to is a function of the verbal environment. Consider the inflection on the verb in t. We have seen that, because that inflection occurs in that place, the corresponding proposition has Psx and Pxj as members. So we *could* conceivably see that inflection as denoting (or, strictly, as denoting2) those properties. When that same inflection occurs in a different environment, e.g. a token of "Plato snores", it denotes a different thing.

I that the first analysis is the preferable one: it is, I think, more natural to see grammatical inflections, and the like, *not* as themselves referring to anything, but as assigning referents to sequences of expressions on the basis of the referents of the components of those sequences.

In effect, we've already seen why this path is preferable to the other. If we taken other path, then we must say that an occurrence of a grammatical inflection denotes a number of different things – e.g. the ending on the verb in t denotes both Psx and Pxj. That inflection is surely not *ambiguous* between those two properties. At the same time, we probably don't want to say that it denotes some *complex* entity consisting of both those properties. What would that entity be? Would it be some set consisting of exactly of those properties? If we agreed to such a proposal, that would unnecessarily complicate our analysis of propositions. So it seems better to say that, strictly speaking, grammatical inflections do not themselves refer, but that they *assign* referent so sequences of expressions on the basis of the referents of the components of

those sequences. At the same time, the facts do not *compel* us to say this: such inflections *can* be treated as contextual referring terms – terms whose referents are a function of the verbal environment in which they occur. The obstacles to taking this path are considerations of simplicity and the like – so far as I can tell, it would not be *wrong* to do so. But for the reasons given a moment ago, my official position is that it is best, from the viewpoint of theory-construction, not to regard grammatical inflections as referring terms.

Some Metaphysical Consequences

Thought separates things that, in reality, are not separable. We think that Smith is a discrete, isolable constituent of the universe. But he is not. Wherever Smith exists, so does a state of affairs. This is the case for a number of reasons. First, nothing is propertyless. If Smith exists, he must have a certain height, weight, and so on. So what we find in reality is never just *Smith*, but Smith's having such and such properties. This is equivalent to saying: what we find is not *Smith*, but always some state of affairs involving Smith. Also, Smith's existence supervenes on the occurrence of innumerable states of affairs: sub-atomic, atomic, molecular, cellular, metabolic, psychological. For modal reasons, Smith cannot be *identified* with the sequence of events on whose existence his own supervenes. But he is plainly inseparable from them: there but for the grace of those lower level existents goes he.

The truth is that Smith is more in the nature of an abstract. He is no separable from the existence of concrete states of affairs than are instances of whiteness or roundness. People can no more exist outside of states of affairs than smiles can exist outside of people (or Cheshire cats).

This suggests that Smith is really a *property* of states of affairs. What is it for Smith to exist in a world. The old picture is that Smith is at the bottom, and properties are heaped on top of him. This picture is reinforced by language, which gives Smith his own, semantically simple expression; and which represents the attribution of properties to Smith as the adding of gratuities to an already self-supporting entity. Presumably this fact about language has its basis in deep cognitive facts. Be all of this as it may, it is the wrong metaphysical model. We find states of affairs first. The basic constituents of a world are displacements of mass-energy, wrinkles in space-time. If those wrinkles are of such and such a type – if those wrinkles have the

right properties – then Smith exists in that world; if not, not. So Smith, I would suggest, is a property of states of affairs.

On similar grounds, I would argue that the things we call "properties" – redness, roundness, and the like – are really properties of properties. Actually, they are, I believe, properties of properties of properties -- 3rd order properties!

First of all, you never find instances of redness in isolation of states of affairs. The same is true of instances of roundness, sweetness, and anything else that we refer to as a "property". What you find is a red (or sweet or round) state of affairs.

What is a state of affairs? Pick some state of affairs. (To facilitate things, choose one that you can see or feel.) Consider the fact that there is a tree in such and such place at such and such time. Let S be that state of affairs. What is it for a world to comprise S? S's existence in a world supervenes on the existence of innumerable displacements of mass-energy – ultimately, on the presence of certain "wrinkles in the quantum" of that world. So that state of affairs exists in a world w if some of the wrinkles in that world's quantum are of the right type, i.e. if those wrinkles have the right properties. So a state of affairs is a property of a property. Instances of redness, tallness, and the like – the things we are used to describing as "properties" – are properties of states of affairs. So they are properties of properties of properties.

Volume II : Cognitive Content
Chapter 17 Introduction to Volume II

This part of the book is going to be quite involved. It will help if I start out by making it absolutely clear what I will be arguing for. So in this chapter, I will give provide a dogmatic statement of the basic views I will be defending. Arguments will be outlined, but not given in their fullness. The arguments proper will be given in subsequent chapters. Since this chapter simply states my views, and doesn't put them in a broad argumentative context, much of what is said here will appear counterintuitive, even alarming. That appearance will vanish in due course.

We must distinguish two quite different doctrines: we must distinguish semantic externalism from *content*-externalism.

I will argue *for* semantic-externalism and *against* content-

externalism.

Content-externalism is roughly this. Let X and Y be two people who differ *at most* in this one respect: the causal origins of X's states differ from the causal origins of Y's mental states. In that case, says the content-externalist, X's mental states may have a different representational content form Y's mental states. X may be thinking about water, while Y is thinking about twin-water. X may be thinking about Mary, while Y is thinking about twin-Mary or perhaps about nothing (Y is hallucinating).

Given two people who are exactly the same, modulo facts about the causal origins of their respective conditions, they may be thinking different things.

Semantic externalism is a parallel, but distinct, doctrine. Let X and Y be two distinct entities – they could be individuals or entire populations – that differ *at most* in this one respect: the causal origins of X's states differ from the causal origins of Y's mental states. In that case, says the semantic externalist, the meanings of X's words may be differ from the meanings of Y's words. X may be referring to water, while Y is referring to Twin-Water. X may be referring to Mary, while Y is referring to twin-Mary or, perhaps, to nothing at all (Y is hallucinating).

There can be no doubt that *semantic*-externalism is correct. Imagine the following. Let w and w* be two worlds that are exactly the same except in this one respect: in w, somebody x uniquely invented the zipper. In w*, there is no such person.

In w, English speakers gather together and say: *if* somebody x uniquely invented the zipper, then let "Julius" refer to x; if there is no such person then "Julius" doesn't refer at all, and sentences of the form "… Julius…" are abortive.

In w*, English speakers gather together, and create the very same semantic rule.

Remember that, with the one qualification, w is exactly like w*. So for any individual in the one world, there is somebody in the other world who is thinking and feeling the exact same things.

It is a simple fact that, in w, "Julius" refers, and that "Julius was either tall or not tall" is true. It is a simple fact that, in w*, "Julius" does not refer and "Julius was either tall or not tall" doesn't encode any proposition, thus being neither true nor false. There are no mysteries here: the situation just described is an artifact of our stipulations.

Suppose that w and w* are exactly the same with this one difference. In w, Smith is seeing Mary, whereas in w*, Smith is having is having a hallucination. Smith-in-w*'s hallucination is qualitatively just like Smith-in-w's veridical perception. With the consent of all his co-linguists, Smith-in-w says: let us refer to *that* person, whoever it is, as "Mary". With the consent of all *his* co-linguists, Smith-in-w* says the exact same thing. In w, "Mary" refers. In w*, it doesn't refer: tokens of "Mary snores", and the like, are abortive.

What our *words* mean – not necessarily what *we* mean – is not entirely up to us. We can impose *some* limits on what our words mean; but some parameter of word meaning is typically fixed by facts about the external world. We can put limits on what "Julius" refers to: we can stipulate that, if somebody uniquely invented the zipper, the "Julius" refers to x. But it is not up to us who that somebody is; it is not up to us whether that somebody exists. So we cannot, by our own fiat, completely nail down the meanings of our own words: we do some of the work; facts about the external world do the rest.

It might seem that, in some cases, this is not so. Suppose I go up to some specific individual and, with the consent of all my fellow English-speakers, I say: let "Julius" refer to *this* person. Let Ralph be the person in question. Here, it seems, I *have* completely pinned down who "Julius" refers to. But this is not so. In effect, we saw why a moment ago. Let w* be some world where the very same thing happens, and where your mental states are exactly like the ones you have in this world (modulo facts about their external causes), but where the individual you dub "Julius" isn't Ralph, but is merely some android that looks exactly like him. Let R-Z745 be that android. In w*, "Julius" refers to R-Z745; in our world "Julius" refers to Ralph.

What our words mean does *not* supervene on facts about our conditions; what our words mean *does* supervene on facts bout our conditions *plus* facts about the external causes of those conditions.

Content-externalism and semantic-externalism are entirely different doctrines; and they have systematically and pervasively been confused.

What my *words* mean is not always what *I* mean. In fact, when we are dealing with words that refer to externalism objects, it is, I will argue, impossible – epistemically and conceptually impossible – that there should be a perfect fit between speaker-meaning and word-meaning.

An object in the external world can never *just* be grasped; it must

always be grasped by way of a uniquely individuating description. This is a consequence of the *predicational* nature of sense-perception. One doesn't see a thing; one sees a state of affairs involving a thing. The content of one's perception is not *Fido.* It is *a dog (or entity) having such and such shape, moving about in such an such a setting.*…The content of one's perceptions is, at least in part, *existential* (*there is some creature x that is moving about in such and such a way*…) A consequence of this (though not an obvious one) is that the information that a given person *associates* with a token "Julius snored" or "Socrates smoked" is a Russellian-existence claim, even though such a claim is not the literal meaning of such a token.

The failure to distinguish what our words mean from what *we* are thinking is one of the cornerstones of content-externalism. Obviously *if* there were a tight fit between word-meaning and thought, *then* the truth of semantic externalism would entail the truth of content-externalism. But there isn't, so it doesn't.

The other cornerstone of content-externalism is a failure to distinguish the property of being *made true* by something from the property of being *about* something. A token of the sentence "Sally met a man today" is made true by the fact that Sally met Harry or Frank or Charlie. But that token is not *about* any of those individuals. We will argue that the external world is given to us through existence claims: sense-perceptions are existential. Suppose that two people have perceptions that encode the very same existence claim. It may still be that what makes the one existence claim *true* differs from what makes the other existence claim true. What *makes true* Smith's thoughts might be the liquidity of *water*; what makes true Twin-Smith's thoughts is the liquidity of twin-water. But that doesn't mean that Smith's thoughts have a different content from Twin-Smith's thoughts. In w, Sally meets Frank. She thinks: *I met a man today.* Her thought is made true by the fact that she met Frank. In w*, Sally meets Pete. She thinks: *I met a man today.* Her thought is made true by the fact that she met Pete. The thought Sally has in w identical with the thought she has in w*. What differs is the *truth*-maker, not the representational content. External objects are given to us through existence-claims. Smith-in-w's thoughts, and perceptions, are *made true* by the fact that water quenches thirst; Smith-in-w*'s thoughts are *made true* by the fact that twin-water quenches thirst. But the *content* of Smith-in-w's thoughts is identical with the content of Smith-in-w*'s mental states.

Ultimately, both of the root-problems with externalism coalesce into a single problem: the failure to distinguish perceptual information from *meta*-perceptual information. Unless I am looking through an electron microscope, there is nothing in my perception itself that tells me that I am seeing H_2O as opposed to ABC. The information *embodied* in the sense-perception obviously has nothing to do with the chemical structure of anything. Later on, in light of new sense-perceptions and in light of my theoretical working over of those perceptions, I may judge that what I was seeing on that occasion was H_2O. But the information embodied in the perception itself – what the perception tells me – has not a whit to do with the chemical structure of anything. The information given me by that perception itself – as opposed to the information that I read back into it on a later date, in light of other sense-perceptions and various scientific constructions -- is as consistent with my seeing something that turns out to be ABC or XYZ as it is with my seeing something that is H_2O. If we focus only on what that perception itself tells me, there is no representational difference between that perception my doppelganger's sense-perception. So it is spurious to say that H_2O is part of the representational content of my perception, whereas ABC is part of the representational content of doppelganger's sense-perception. My perception tells me *nothing* that my doppelganger's perception does not tell him, and *vice versa.* Given only the information bequeathed me by that very perception, there is no inference I can make that my doppelganger cannot, with equal rationality, make on the basis of his perception, and *vice versa.* If there *were* a difference in the representational content of our perceptions, that would show up in a difference in what we can conclude on the basis of them. For differences in content just *are* differences in what is inferentially licensed (a controversial claim: but one we will defend). There is nothing that I can rationally believe on the basis of my perception that my doppelganger cannot also, with equally rationality, believe. Differences in rationality appear, if at all, only when we consider perceptions *additional* to the two just mentioned – only when we consider ulterior beliefs: beliefs not encoded in the perceptions themselves. If my doppelganger does experiments (or hears about experiments) suggesting that what he saw was ABC, while I do (or hear about) experiments suggesting that what I saw was H_2O, then what I can rationally believe will start to diverge from what my doppelganger can rationally believe. But nothing in the perception itself warrants such a

divergence: and, since the representational content of a mental state is inseparable from the inferential liaisons of that state, this means that what my perception tells me is *not* different from what my doppelganger's perception tells – even though (see above) those perceptions have different *truth-makers.*

Chapter 18 Extensionality and cognitive content

We've just said what we are going to argue for. Providing those arguments will involve our entering a dark and chaotic wilderness of epistemological dispute. Let us begin our foray into that wilderness by discussing a problem that, at least on the surface, is narrowly semantic: the problem of *extensionality.* The semantic points we make will provide us with a much needed beacon in our odyssey through the obscurities of epistemology.

We've already seen some reason to think it impossible to change meaning by replacing a referring term with a co-referring term – some reason to think that all contexts are *extensional.* This is sometimes known as the doctrine of extensionality. But that is famously problematic, as it seems to be subject to obvious counterexamples:

(a) "John believes that the inventor of bifocals snored" (true)
(b) "John believes that the first postmaster general snored." (false)

(1) "Necessarily, the inventor of bifocals snored is identical with the inventor of bifocals." (true)
(2) "Necessarily, the inventor of bifocals snored is identical with the first postmaster general." (false)

There is no denying that what (a) and (1) communicate is different from what (b) and (2) communicate. But for many reasons – some of which we've seen, some of which we will soon see -- (a) and (b), and (1) and (2), probably coincide in literal meaning.

Right now I wish to show that, if we keep in mind a few basic points about sense-perception, and we also take care to distinguish literal from

communicated meaning, we can make these problems vanish; we can vindicate extensionality, and – what is more important – we can explain exactly why there *appear* to be counter-examples to it.

In some cases, referring terms are defined ostensively. I point to a celestial body and I say: "*that* [pointing to a certain entity] is named ‘Hesperus’".

On the face of it, referring terms are not always defined in this way. Sometimes they are defined *by description.* I don’t *point* to anything and say "that is ‘Hesperus’". Instead, I say:

(11) "Hesperus is the first celestial body to appear in the evening sky."

Here, of course, I am defining "Hesperus" in the sense that I am giving its referent.

When I *define* "Hesperus" for someone by using a sentence like (11), what I am really saying is this:

(12) "’Hesperus" is the name of the first celestial body to appear in the evening sky."

I should make a certain nuance as clear as possible. Suppose Smith is *already* acquainted with Hesperus; suppose he is an astronaut who has actually set foot on Hesperus, and is acquainted with it in that way. Further, suppose he knows that thing to be called "Hesperus". Of course, Smith might not know that Hesperus is the first celestial body to appear in Earth’s evening sky. So if I uttered (11) to him, what I would be saying would *not* be equivalent to (12); I would not make making a metalinguistic statement; I would be making a garden-variety objectual statement.

But suppose Jones does not know what "Hesperus" means, and I utter (11) to him. In *that* case, (11) is tantamount to (12); in *that* case – in the case where I am telling him what "Hesperus" refers to -- what I am saying is really (12).

For reasons we’ve seen,

"…the phi…"

communicates the proposition:

exactly one thing O has phi and…O…

So (12) *communicates* the proposition:

(13) Something O is uniquely a first celestial body to appear in the evening and "Hesperus" names O.

Let us suppose that "Hesperus" has, in fact, been defined for you (in the reference-fixing, not the meaning-giving, sense) by (xi), or some other similar sentence. Since (12) communicates (though it doesn't literally mean) (13), a consequence is that sentences of the form

"…Hesperus…"

communicate (not semantically encode) propositions of the form:

Something O is uniquely a first celestial body to appear in the evening sky and "Hesperus" names O and…O…

So what

(14) "Hesperus is lovely"

communicates to you is roughly this.

(15) There is some object O such that O is uniquely a first celestial body to appear in the evening sky and "Hesperus" names O, and O is lovely.

Thus, *part* of what (14) communicates to you is:

(16) There is some object O such that O is uniquely a first celestial body to appear in the evening sky and O is lovely.

Suppose Jones does not know what "Phosphorous" means. You tell him:

(17) "Phosphorous is the last celestial body to disappear from the morning sky".

Under this circumstance, (17) amounts to:

(18) "'Phosphorous" names the last celestial body to disappear from the morning sky".

Again, we must be very careful about one thing. (17) doesn't *necessarily* communicate the same thing as (17). Once again, suppose that Smith is an astronaut who is already acquainted with Phosphorous (Hesperus), and is therefore acquainted with it. And suppose he also knows that it is sometimes called "Phosphorous". But Smith does not know that that thing is the last celestial body to disappear from Earth's morning sky". In that case, (17) will not have the same import at (18).

But if Jones simply doesn't know what "Phosphorous" refers to, and you utter (17) to him, you are really tell him something metalinguistic; you are telling him what "Phosphorous" refers to; you are, in effect, uttering (18) to him.

For reasons exactly analogous to those given earlier, it follows that, where Jones is concerned, (17) communicates the proposition:

(19) Something O is uniquely a last celestial body to disappear from the evening sky and "Phosphorous" names O.

And, again for reasons exactly similar to those already given,

(20) "Phosphorous is lovely"

will *communicate*---though, as we know from Kripke, it won't

semantically encode---the proposition:

(21) Something O is uniquely a last celestial body to disappear from the evening sky and "Phosphorous" names O and O is lovely.

So, under the circumstances described, part of what is communicated by (20) is:

(22) Something O is uniquely a last celestial body to disappear from the evening sky and O is lovely.

So, under the circumstances described, what is communicated to Jones by "Hesperus is lovely" is very different from what is communicated to Jones by "Phosphorous is lovely". The proposition communicated by the first does not entail, and is not entailed by, the proposition, communicated by the second.

There are epistemically possible worlds where

(15) There is some object O such that O is uniquely a first celestial body to appear in the evening sky and "Hesperus" names O, and O is lovely.

is true and

(21) Something O is uniquely a last celestial body to disappear from the evening sky and "Phosphorous" names O and O is lovely.

is false, and *vice versa.*

And yet "Hesperus" is synonymous with "Phosphorous". Indeed, we've built that fact into our story.

"…Hesperus…" does not *have* to have a different cognitive value from "…Phosphorous…" Once again, consider cosmonaut Smith. He has actually set foot on Hesperus (Venus/Phosphorous), and he also knows that that thing is sometimes called "Phosphorous" and is sometimes called

"Hesperus". So, for Smith, "Phosphorous is lovely" and "Hesperus is lovely" will have (nearly enough) the same cognitive value.

But most people learn what "Hesperus" and "Phosphorus" through non-ostensive (descriptive) definitions. And the definitions used typically exploit *different* descriptions. For reasons we've already seen, this results in a difference in cognitive value, at least for those people, between "… Hesperus…" and "…Phosphorous…"

Obviously this point generalizes. It applies to the possible difference in cognitive value between "…Cicero…" and "…Tully…" . Indeed, for *any* two co-referring names E and E* whose meanings on learns through non-ostensive definitions, our analysis explains the possibly difference in cognitive value between "…E…" and "…E*..."

Ostensive definition and extensionality

The problem now is extending our remarks to cases where one learns word-meaning through *ostensive* definition. Suppose I learn what "Hesperus" means through, and also what "Phosphorous" means, through *ostensive* definitions. It is quite possible that, under that circumstance, "… Hesperus…" will have a very different cognitive value from "… Phosphorous…" We need to deal with this fact.

Let us suppose that, during late morning, I am talking with Brown as, and he asks me what "Hesperus" means or refers to. I point to a certain object, and say

(22) "*that* [pointing to the last celestial body, other than the sun, to remain in the sky] is Hesperus".

Later, during the onset of dawn, I am again talking to Brown. He asks me what "Phosphorous" refers to. I point to a certain object, and say:

(23) "*that* [pointing to the first celestial body to appear in the sky] is Hesperus".

Under these circumstances, "…Hesperus…" will communicate a very different proposition from "…Phosphorous…", i.e. those sentences will

have different cognitive values.

This fact fits in with what we've already seen. But to see how it fits in, we must be very clear on a certain epistemological fact.

The predicational nature of sense-perception

Suppose you look at a physical object. It could be any physical object – a dog, a vase, a planet. Let O be that object, whatsoever you choose it to be.

When you look at O, what is the content of your perception? What is your perception really telling you?

You don't *just* see O. You see (say) an object with a certain shape, in a certain place, moving in a certain way, sitting on top of some other object, and so on. You don't *just* see O; you see a state of affairs involving O.

For the sake of discussion, suppose that O is a red apple, with a peculiar shape, and you are looking at O. Let P be the perception, or series of perceptions, in question. A blind friend of yours asks you tell him what you see. To give the content of P, you say something like:

(ES) "there is a red apple; it doesn't have quite the normal apple-shape, rather it is shaped like […]; it is on top of a hexagonal table […]"

The content of your perception is *existential*; it is given, at least in part, by an *existence*-claim:

(EP) *there is an object x such that x has a certain peculiar shape and x is a hexagonal table y…*

The content of a perception is never just an object; it is a state of affairs involving an object. That state of affairs is given by existential information.

That existential information may itself be object-involving. When I am describing P to my blind friend, I may say something like:

(EP*) "There is an apple in *Fred*'s hand whose shape is […], and it is on a hexagonal table on which *Tim* is sitting […]"

But what this means is that P is a perception not just of O, but of Fred and Tim as well. Just as the content of my perception of O is given by an existence claim, so is my perception of Fred and Tim. I don't *just* see Tim; I see an entity with a certain shape, sitting in a certain position, in a certain location…So what I see is given by an existence claim:

there is an entity x whose shape is […], whose seating-position is […], and is located…

Whenever you sense-perceive a thing, the content of your perception is given by an existence claim. When you see Fido, you don't *just* see a dog; you see a thing with four legs, with fur of a certain color, moving about in a certain way. The content of your perception is existential: *there is a thing with four legs and fur of a certain color…*

In light of this point, let us revisit the Hesperus-Phosphorous case. Jones wants to know what "Hesperus" refers to. I point to a certain body in the morning sky, and say: *That* is named "Hesperus". Obviously my ostensive definition works only because Jones is having a sense-perception of the thing indicated by the "that". If Jones were blind, the definition would be useless. What is the content of Jones' perception? That content is existential. Jones is, of course, seeing Hesperus. But he is not *just* seeing Hesperus. He is seeing *an* object with a certain luminosity, a certain shape, a certain relative position in a sky of having a certain distinctive hue, and also having the property of being alone in said sky.…So the content of his perception is, at least in part, existential:

(*) There is a certain object x such that x has […] shape and […] luminosity and […] and being alone in the morning sky and […]

So, when I was pointing out Hesperus to Jones, and saying that it was called "Hesperus", (*) is the perception Jones was having; it is the perception that gave my definition any force. Consequently, the import, for

Jones, of my definition is this:

(**)*There is a certain object x such that x is alone in the morning sky and x has […] shape and […] luminosity and […]* and "Hesperus" names x.

For reasons we have already seen, this means that, for Jones, the cognitive significance of

(3) "Hesperus is lovely"

is

(4) *There is a certain object x such that x is alone n the morning sky, and x has […] shape and […] luminosity and […]* and "Hesperus" names x and x is lovely.

Remember how "Phosphorous" was defined for Jones. I pointed to the an object in the evening sky, the only one yet to appear, and said: *That* is named "Phosphorous". For exact analogues of the reasons just given, the import, to Jones, of my definition was:

(5) There is a certain object y such that x is alone in the evening sky and y has […] shape and […] luminosity and […] and "Phosphorous" names y.

And, for reasons exactly similar to those just stated, a consequence is that, to Jones,

(6) "Phosphorous is lovely"

communicates:

(7) *There is a certain object y such that x is alone in the evening sky and y has […] shape and […] luminosity and […]* and "Phosphorous" names

y and y is lovely.

(3) communicates (4), and (6) communicates (7).

(4) does not entail (7), and (7) does not entail (4).

There are epistemically possible worlds where (4) is true and (7) is false, and there are epistemically possible worlds where (7) is true and (4) is false.

In fact, there are *metaphysically* possible worlds where (4) is true and (7) is false, and there are *metaphysically* possible worlds where (7) is true and (4) is false.

Taking stock

We know that "Hesperus" and "Phosphorous" are just labels, and that each has only one thing – Venus – for its semantic content. We also know that

(3) "Hesperus is lovely"

can have a very different cognitive value from:

(6) "Phosphorous is lovely"

When we are dealing with a case where "Hesperus" and "Phosphorous" have been defined non-ostensively – that is, they have been defined descriptively – it is easy to produce an explanation for this difference in cognitive value. The explanation turns on the fact that descriptive definition is really *existential* definition. To say:

"Hesperus" names the last celestial body to disappear from morning sky

is really to say

Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x.

An immediate consequence is that, if "Hesperus" has been thus defined for you, what "Hesperus is lovely" will communicate to you will be an existence claim:

(#) Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x and x is lovely.

For exactly similar reasons, if "Phosphorous" is defined non-ostensively, what "Phosphorous is lovely" will communicate to you will be some other existence claim, quite possibly one that doesn't entail, and isn't entailed by (#).

Extending this analysis to ostensively defined terms

"Hesperus" and "Phosphorous" can be defined either descriptively or ostensively. We've seen how, when they are defined descriptively (non-ostensively), "…Hesperus…" and can differ in cognitive value from, "… Phosphorous…", even though they are perfectly synonymous. So we have accounted for *some* of the cases where replacing referring terms with co-referring terms seems not to preserve literal meaning. But we have not yet accounted for all such cases: after all, proper names are sometimes defined ostensively. Now we must show how the things we've just said about descriptive (non-ostensive) definition can be made to apply to cases where "Hesperus" and "Phosphorous" are defined, not descriptively, but ostensively. There is no problem here.

As we've seen, all perception is "existential" (or "predicational" as Barry Stroud put it). When you see something, the content of your perception is given by an existence claim. The existence claim involved in your taking in an ostensive definition of "Hesperus" may be very different from the one involved in your taking in an ostensive definition of "Phosphorous"; in that case, (3) will communicate one existence claim, and (6) will communicate a different one; and, quite possibly, neither will entail the other.

Rationality and literal meaning

(3) and (6) encode the very same proposition. But one can assent to (3) and *deny* (6) without being guilty of irrationality. This seems like a paradox. It seems as though assenting to (3) and denying (6) ought to amount to affirming and denying the very same proposition. What has gone wrong?

We've seen what has gone wrong. At the level of literal meaning, (3) and (6) both encode the same proposition: *Venus is lovely*. But what (3) *communicates* is some existence claim E; and what (6) communicates is some other existence claim E*. And E neither entails, nor is entailed by, E*. Rationality is to be measured not in terms of literal meaning, but in terms of what one *grasps* – what is communicated to one.

Given the right circumstances, any sentence, no matter what its literal meaning, can convey anything. Suppose a person who you know to be extremely violent, and who you know to carrying a fire-arm, says

(*) "I think it would be a good idea if you left immediately".

The literal meaning of (*) is some claim about what a certain person thinks. The communicated meaning is:

(**) I am going to shoot you if you do not leave immediately.

It would be irrational of me to base my decisions on the literal meaning, and not the communicated meaning, of (*). The yardstick of rationality is communicated, not literal, meaning.

(3) and (6) have the same literal meaning. But they can communicate different things. When Jones assents to (3), what he is *really* assenting to assenting to is (4). If Jones should deny (6), he is, in effect, denying (7). The conjunction formed of (4) and the negation of (7) does not have the form *P and not P;* that conjunction is not, by anyone's lights, self-contradictory. That is why Jones can *rationally* assent to (3) and also to the negation of (6).

Kripke's Pierre-paradox

Kripke (1977) produced an argument that is meant to undermine the

view that, at the level of semantics, names have senses. In my view, that argument is cogent and devastating.

That same argument *seems* to warrant the view that, in at least some cases, one can assent to a proposition P *and* its negation *without being irrational.* In light of the points made a moment ago, we will see that this view is unwarranted. Without qualification, it is *always* irrational to assent to *P and not P*.

In effect, we've already seen why this is so; but we need to adapt the points earlier made to fit Kripke's argument.

The argument is given by the following thought-experiment. Pierre doesn't speak a word of English. Pierre does not know that "London" is the English translation of "Londres";

Pierre knows that there is a city in English called "Londres", in French; and he is has heard from reliable sources it is a lovely place.

In consequence of this, Pierre rationally assents to:

(*) "Londres est jolie".

One day, Pierre is kidnapped and kept in a rundown section of London. He learns English from scratch; he does *not* learn it by having French locutions translated into English. He learns that he is living in a place called "London". All the data that Pierre has suggest that he is living in an unattractive city. So Pierre rationally assents to the sentence:

(**) "London is *not* pretty".

The proposition encoded in (**) is the negation of the proposition encoded I (*). Pierre rationally assents to both sentences.

Many have taken this to show that one *can* rationally assent to a proposition and its negation. So many have taken Kripke's thought-experiment to require a rejection of an extremely reasonable and millennia-old conception of rationality.

They have said that whether it is rational to assent to *P and not P* depends on the "guises" through which one grasps those propositions. So Smith can believe: *Venus is lovely*, and also believe: *Venus is not lovely.* Smith isn't necessarily guilty of irrationality. If he believes those propositions

under the right "guises", Smith is still rational. To be irrational, there must be some *one* "guise" G such that Smith believes *Venus is lovely* under G and also believes *Venus is not lovely* under G.

I don't think we have to go such lengths. The traditional view is: if you believe *P and not P*, you are irrational. I don't think it is necessary to revise this view, even in light of Kripke's point.

How does Pierre learn the meaning of "Londres"? He learns it either descriptively or denotatively. For our purposes, it doesn't matter which. In either case, he learns it through some existential proposition, something along the lines of:

(F) *there is some beautiful city x in England and "Londres" names x*

As we have seen, a consequence is that the cognitive value, for Pierre, of

(*) "Londres est jolie"

is

(LEJ) *there is some beautiful city x in England and "Londres" names x and x is pretty.*

How does Pierre learning the meaning of "London". He seems to learn it ostensively: *This* place is called "London". As we have seen, this means that he is, in effect, learning it through an existence claim:

(U) *There is some city y that I am now occupying such y is ugly and "London" names y.*

A consequence is that what "London is not pretty" communicates to Pierre is:

(LNP) *There is some city y that I am now occupying such y is ugly and "London" names y and y is NOT pretty.*

(LEJ) does not entail, and is not entailed by, (LNP).

(LEJ) is what is communicated to Pierre by (*) "Londres est jolie".

(LNP) is what is communicated to Pierre by (**) "London is not pretty".

So the proposition that Pierre assents to, in assenting to (*), is (LEJ).

And the proposition that Pierre assents to, in assenting to (**), is (LNP).

By anyone's lights, *LEJ and LNP* is not a contradiction.

That is why Pierre is not guilty of irrationality in assenting to (*) and also to (**).

Pierre's being rational is thus perfectly consistent with the traditional view that, for any proposition P, one is unconditionally irrational if one assents to *P and not P*.

This analysis has yet further consequences. The sentence

(a) "Hesperus is Hesperus"

obviously encodes an analytic proposition. Everybody agrees with this.

These days, everybody agrees that

(b) "Hesperus is Phosphorous"

encodes a *necessary* proposition, but not an analytic one. Why is it held to be necessary? If x *is* y, then x couldn't possibly *not* have been y; a thing cannot fail to be identical with itself; and since y *is* x, then x's not being identical with y is the same as x's not being identical with itself. So given that x is identical with y, x *must* be identical with y.

But surely, we are told, (**) is not *analytic*; the proposition is encodes is empirical it is not in the same category as a truism like *Hesperus is Hesperus.*

There is a major problem with this view. "Hesperus" and "Phosphorous" merely label a certain entity; they are not Russellian quantifiers. The reason that (**) expresses a *necessary* proposition is, specifically, that "Hesperus" and "Phosphorous" are just labels, and not

Fregean sense-bearing terms or Russellian-quantifiers. But, by that token, (**) must encode the proposition

(c) *Hesperus is Hesperus.*

(c) certainly looks analytic.

At the same time, it is pretty clear that one can speak English and understand (**), and assent to it, without being irrational – without believing that Hesperus is *not* identical with Hesperus.

People have dealt with this last fact by saying that:

> *(**) expresses a proposition that is metaphysically necessary, but epistemically contingent. Given that Hesperus is identical with Phosphorous, it follows that there is no metaphysically possible world where Hesperus is distinct from Phosphorous. But given only the information embodied in (**), one has no way of knowing whether it is true or false. So the proposition encoded in (**) is epistemically contingent: it could turn out either way.*

But we've already seen the problem here. The very reason that (**) expresses a necessary proposition is that "Hesperus" and "Phosphorous" are labels, not generalized quantifiers. But that very fact about them makes (**) express (c), and thus be synonymous with (*).

This can all be sorted out. First of all, we must distinguish *sentences* from propositions. And we must also distinguish literal meaning from communicated meaning (cognitive significance).

(*) and (**) have the same literal meaning. But they may *communicate* very different propositions. Let us be more precise about this. Once again, consider Jones. For him

(3) "Hesperus is lovely"

communicates a proposition of the form:

(4) Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x and x is lovely.

In general, for Jones,

(8) "…Hesperus…"

communicates a proposition of the form

(8) Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x and…x…

And, for Jones,

(6) "Phosphorous is lovely"

communicates a proposition of the form:

(4) Something x is uniquely a first celestial body to appear in the morning sky and "Phosphorous" names x and x is lovely.

In general, for Jones,

(9) "…Phosphorous…"

communicates something of the form:

(10) Something x is uniquely a first celestial body to appear in the morning sky and "Phosphorous" names x and…x…

It follows that, for Jones,

(**) "Hesperus is Phosphorous"

communicates:

(11) Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x, and something x is uniquely a first celestial body to appear in the morning sky and "Phosphorous" names x and x=y.

(11), of course, is a *contingent* proposition: it is contingent in every sense; there are metaphysically (not just logically) possible worlds where (11) is false. Everyone will grant this.

In light of this, consider the sentence:

(***) "necessarily, Hesperus is Phosphorous".

Take the word "necessarily" to denote analytic (or "a priori") necessity: the kind of necessity characteristic of "triangles have three sides". For the moment, let us forget about the "a posteriori" kind supposedly characteristic of "water is H_2O".

In (***), sentence falling within the scope of the "necessarily" *communicates* (11). The "necessarily" can be given different degrees of scope *within* (11). It can be given wide-scope or narrow-scope. Let us consider the *wid*est-scope reading and the narrow*est*-scope reading.

The narrow*est*-scope reading is:

(12) Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x, and something x is uniquely a first celestial body to appear in the morning sky and "Phosphorous" names x and **necessarily**: x=y.

The widest-scope reading is:

(13) **Necessarily**: Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x, and something x is uniquely a first celestial body to appear in the morning sky and "Phosphorous" names x and x=y.

(13) is false. This fact corresponds to the fact that (**) is "epistemically contingent".

(12) is true. This fact corresponds to the fact that (**) is "metaphysically necessary".

When we say that (**) is "metaphysically necessary, but epistemically contingent", that is a very sloppy way of saying: (**) encodes a proposition that is analytic - -"necessary" in the traditional sense – but *empirical work is needed to figure that out.*

The proposition semantically encoded in (***) is

(14) necessarily: Hesperus is Hesperus.

(14) is analytic. After all, for any object O, the proposition:

(15) *O is identical with O*

is analytic. In (12), the proposition falling within the scope of the "necessarily" is one of the same form as (15). The kind of necessity involved in (12) is analytic necessity; it isn't any other kind; it isn't some kind that was undiscovered prior to Kripke.

(12) is a proposition, not an expression. The values of its variables are objects, not expressions. Those values are individuals: not expressions, not Fregean senses, not Russellian quantifiers, not inscriptions or noises.

(12) is also a *true* proposition. This means that, for some value of x

and for some value of y, if we replace x and y in (12) with those values, the result is a true singular proposition. Let Alpha and Beta be *objects* – not expressions – such that, when we replace the x in (12) with Alpha, and the y in (12) with Beta, the result is a true singular proposition:

(16) Alpha is uniquely a last celestial body to disappear from morning sky and "Hesperus" names Alpha, and Beta is uniquely a first celestial body to appear in the morning sky and "Phosphorous" names Beta; and **necessarily**: Alpha=Beta.

Notice that (16) is a conjunction. Let us focus on the last conjunct.

(17) necessarily: Alpha=Beta.

Obviously Alpha and Beta must be the very same object. Otherwise (16) is an immediate failure. So (17) is identical with the proposition:

(18) necessarily: Alpha is Alpha.

So the "necessarily" in (12) is plain old analytic necessity. It is not "metaphysical" necessity; or, if it is, then metaphysical necessity is the same thing as analytic necessity, at least in this case.

Let us take stock. We are told that what is meant by:

(**) "Hesperus is Phosphorous"

is "necessary" but "a posteriori"; sometimes we say that it is "metaphysically necessary" but "logically [epistemically] contingent".

But this simply isn’t true. The proposition encoded in (**) is:

(c) *Hesperus is Hesperus.*

(c) is analytic.

The information *through* which we grasp the literal meaning of (**) is not analytic, at least not typically. That information is given by an existence claim, something along the lines of

(11) Something x is uniquely a last celestial body to disappear from morning sky and "Hesperus" names x, and something x is uniquely a first celestial body to appear in the morning sky and "Phosphorous" names x and x=y.

(11) is quite contingent. If we put a necessarily-operator into (11), and give it *narrowest* scope, then what results is either necessarily true or necessarily false. But the necessary truth will be an analytic necessity, and the necessary falsity will be an analytic falsity.

If we give a necessity-operator widest-scope, what results is simply false.

When we say that (11) is true in all metaphysically possible worlds, what we are saying, in so far as we are being coherent, is that the narrowest-scope reading is true.

When we say that (11) is false in some epistemically possible worlds, what we are saying is that the widest-scope reading is false.

When we say that sentence S expresses a "necessary a posteriori" truth – when we say that S expresses a truth that "holds in all possible worlds, but is not analytic or logically true" -- what we are really saying, in so far as we are not just wrong, is: the literal meaning of S is analytic, but empirical work is needed to figure out that fact.

Chapter 19 Externalism and self-knowledge

These points give us some leverage on a problem that has exercised the finest minds of our discipline for the last thirty years.

Imagine a planet qualitatively identical with ours, but many light-

years away. Let w* be that world. For each object x in our world, there is an object x* in w* that is an atom for atom duplicate of x. In our world, Smith knows Jones. In w*, Smith* knows Jones*. In our world, Smith thinks about Hesperus. In w*, Smith* thinks about Hesperus*.

(To simplify things, let us suppose that, because of the arrangement of the universe, light rays from Hesperus cannot reach w*, and light rays from Hesperus* cannot reach our world.)

When we leave aside the origins of Smith's bodily and mental condition, and when we leave aside the origins of Smith*'s bodily and mental condition, we find that Smith and Smith* are qualitatively identical. We might say this: Smith and Smith* have the same "intrinsic properties". (This is just our short-hand way of saying that, if we want to find to a qualitative difference between Smith and Smith*, we have to look to the origins of their respective conditions: no difference will show up if, leaving aside facts about origination, we consider only the conditions themselves. I am not trying to give an analysis of "intrinsic property"; I am using this expression in a purely stipulative way.)

But Smith has a concept of Hesperus not of Hesperus*, whereas Smith* has a concept of Hesperus *, and not of Hesperus.

Why this difference? Because Smith has a certain connection to Hesperus that he (Smith) doesn't have to Hesperus*, and Smith* has a certain connection to Hesperus* that he doesn't have to Hesperus.

The moral is this. At least, part of one's concept of a thing is a causal connection to that thing. The reason Smith is thinking of Hesperus, and not of Hesperus *, has nothing to do with Smith's intrinsic properties. After all, Smith and Smith* are qualitatively identical, at least in respect of intrinsic properties. The difference lies in the causes of their respective conditions. Smith's states are caused by Hesperus; Smith*'s states are caused by Hesperus*.

So, in at least some cases, one's concept of a thing consists, at least in part, in a causal connection between oneself and that thing.[106] Let us refer to this doctrine as "content-externalism".

A corollary of content-externalism is that two subjects s and s* can be qualitatively identical (leaving aside facts about the origins of their respective conditions) and yet have different concepts and, thus, think different things.

Externalism: some prima facie *problems*

Obviously there is much to be said for externalism; the arguments for it seem quite cogent. But externalism leads to some nasty problems.

From Smith's viewpoint, it is epistemically possible that he is a disembodied spirit and that his "perceptions" of the physical world are all hallucinations. I am not saying that these things are likely to be true or that Smith thinks they are likely to be true. But if, in a moment of skepticism, Smith were to hypothesize that all his "perceptions" were hallucinations induced by an evil Demon, that hypothesis would fit the data at Smith's disposal as well as the hypothesis that his perceptions were veridical, and thus caused by physical objects.

Given all of this, suppose Smith is consciously thinking to himself:

(A) *Hesperus is lovely.*

Further, suppose that Smith goes on to think:

(B) I *know* with incorrigible certainty that I am thinking that Hesperus is lovely.

Obviously Smith may be wrong to think (A). But (B) seems to be in a different category. If Smith thinks that he is thinking (A), then he must be thinking (A). In fact, if he thinks that he is thinking (A), that right there *makes* it true that he is thinking (A). The thought that Hesperus is lovely is a constituent any thought to the effect that one is thinking that Hesperus is lovely. So (B) guarantees its own truth.[107]

In any case, it seems that a certain incorrigibility attaches to (B) that does *not* attach to (A).

Suppose Smith is a brain in a vat, or the dupe of an evil demon. Even so, *if* Smith thinks (B), he is going to be right. The thought is self-fulfilling.

Now we come face to face with a nasty dilemma. If externalism is right, then Smith simply cannot be thinking about Hesperus at all *unless* there is an external world; for Smith's Hesperus-thought has as a constituent a casual connection mediating between Hesperus and Smith.

The right kind of causal connection is lacking. So Smith is shooting

a blank. He isn't thinking about Hesperus at all.

If there is no Hesperus, then Smith cannot be having a Hesperus-thought. A Hesperus-thought requires that there be a causal connection between the thinker and *Hesperus.* No Hesperus, no Hesperus-thought.

At the same time, Smith can*not* know with any kind of incorrigibility that there is an external world or, therefore, that there is a Hesperus. At the same time, there are no Hesperus-thoughts unless there is an external world and, in particular, a Hesperus.

Yet Smith *can* know with absolute certainty that he is *thinking* that Hesperus is lovely. In fact, if he thinks as much, he makes it true. If he *thinks* that he is thinking that Hesperus is lovely, then *ipso facto* he is thinking that Hesperus is lovely.[108]

Given that Hesperus-thoughts presuppose the existence of an external world, and of a Hesperus, the fact that Smith has this incorrigible knowledge seems to constitute irrefragable proof that there is an external world and, in particular, a Hesperus.

So if Smith *thinks* that he is thinking that Hesperus is lovely, then has provided himself with irrefragable proof – proof in the strictest conceivable sense -- that there is an external world and, in particular, a Hesperus. Smith knows with incorrigible certainty that:

(*) I [Smith] am thinking that Hesperus is lovely.

(*) entails that:

(**) Hesperus, and therefore the external world, exist.

So, on the basis of (*) Smith can deduce that there is a Hesperus and an external world. Thus, on the basis of a proposition that Smith knows, with incorrigible certainty, to be true, Smith can deduce that there is a Hesperus, where "deduce" is being used in the strictest sense.

By exactly similar reasoning, Smith can deduce, in the strictest sense, that there is a Ted Kennedy, a George Bush, an Alpha Centauri, and so forth. Externalism gives us incorrigible knowledge of the external world.

Externalism makes the non-existence of Hesperus be *logically inconsistent* with what is going in our private Cartesian theatres.[109]

The problem with this picture

But, plainly, this result is false. Whatever is going on Smith's mind, those goings on are logically consistent with there being no external world and no Hesperus; indeed, they are not only logically, but also nomically, consistent with there being no Hesperus (and no Ted Kennedy and not George Bush…). It is nomically possible that the stimulations which lead to Smith's condition should have been caused by something other than Hesperus.

This must be understood aright. Smith may have knowledge of an external world and, indeed, of Hesperus. (I have no doubt that we *do* know a great deal about the external world.) But surely what is going in Smith's Cartesian theatre does not *logically* necessitate that there be a Hesperus. Let P be any proposition that Smith knows with incorrigible certainty to be true. P does not *entail* that Hesperus exists. P may probabilify the existence of Hesperus, but not entail it.

Solving the problem

This is a nasty dilemma indeed. But everything falls into place when we keep a few platitudes in the forefront of consciousness.

We must distinguish literal meaning from communicated meaning. We must also give due attention to the fact that sense-perception necessarily has an existential component.

Suppose Smith looks up at Hesperus (Phosphorous/Venus). Suppose, also, he has never seen it before and never even heard about it.

What is the content of Smith's perception? Remember, when one sees an object, one doesn't *just* see that object; one sees a state of affairs. One doesn't just see Hesperus. The content of one's perception is given by an existence claim:

(C) there is an object x such that x is alone in the morning sky, having […] shape. and […] luminosity.

Suppose you are there with Smith, as he has this perception, and you say to him, indicating Hesperus: That thing is called "Hesperus".

We've already seen that, under this circumstance, the cognitive significance of

(**) "Hesperus is lovely"

will be some existence claim:

(4) *There is a certain object x such that x is alone n the morning sky, and x has […] shape and […] luminosity and […]* and "Hesperus" names x and x is lovely.

So when we focus on the meaning that (**) has *to Smith* – not its literal, but its cognitive, significance – we find that the planet Hesperus has been existentialized out.

The proposition encoded in (**) is object-involving with respect to Hesperus. But the cognitive significance of (**) is not thus object-involving.

Suppose Smith says:

(D) "I know with incorrigible certainty that I am thinking: that Hesperus is lovely."

The expression ""I know with incorrigible certainty that" denotes an epistemic operator.

What belief is Smith expressing through (D)? It doesn't matter what belief is literally *meant* by (D). For, as we have seen, literal meaning and cognitive meaning may diverge dramatically when we are dealing with sentences containing terms, like "Hesperus", that refer to bits of the external world. What I want to know is: What proposition is Smith expressing through (D)? What is the belief expressed?

We know that when Smith says "Hesperus is lovely", he is expressing (4). Within (4), there are different places to insert an epistemic operator. In particular, it can be given narrowest-scope, or it can be given widest-scope. (For the moment, let us ignore the possibility of giving it intermediate degrees of scope.) So (D) – the result of putting an epistemic

operator in front of (4) – is ambiguous in terms of what it communicates, depending on where the operator is inserted. The widest scope reading is:

(E) **I know with incorrigible certainty that I am thinking that**: *There is a certain object x such that x is alone n the morning sky, and x has […] shape and […] luminosity and […]* and "Hesperus" names x and x is lovely.

The narrowest-scope reading is:

(F): *There is a certain object x such that x is alone n the morning sky, and x has […] shape and […]*
luminosity and […] and "Hesperus" names x and **I know with incorrigible certainty that I am thinking that** : x is lovely.

(E) is not object-involving with respect to any planet. There might be a planet fitting the description, or there might not be. That won't have the slightest affect on whether (E) is true. Given *any* external-world object O, the existence or non-existence of O won't have the slightest affect on whether E is true. For, as we've seen, O will have been existentialized out.

So if we take claims about incorrigible knowledge in the widest-scope way – the de dicto way, the "attributive" way – then they come out true. If we take (D) in the widest-scope way, then Smith is probably quite right. In any case, given that reading, Smith's claim to incorrigibility is not neutralized by the fact that, for all Smith knows, there might not even be a Hesperus. I leave it open whether Smith's claim is neutralized by other, e.g. Freudian, considerations.

But if we take (D) in the narrowest scope way – the referential way, the de re way – then Smith is simply wrong. If we take (D) in that way, then, quite plainly, Smith's certainty about what he thinks *is* only as good as his certainty about the external world.

Let O be some physical object. If you don't know with certainty that

O exists, then you don't know with certainty that the object-involving proposition:

(G) *O is lovely*

exists. For all you know, the proposition you want is no more to be had than a perpetual motion machine or a round square. If you aren't even certain of the very *existence* of:

(G) *O is lovely*

then *a fortiori* you cannot possibly be certain of the very *existence* of the proposition

(H) *I believe that the proposition* O is lovely *is true.*

If you cannot be certain of (H)'s very existence, then you cannot be certain of its *truth.*

So the following two propositions *are* completely incompatible.

(a) content-externalism is true.

(b) For any object-involving proposition P, I can know with incorrigible certainty that I am thinking (or believing) P.

If you know with Cartesian certainty that you believe P, then P cannot be object-involving. What you *can* know with such certainty is that you are thinking some object-*in*dependent proposition: some proposition like (E).

Propositions like (E) can be true.[110] Propositions like (F) are always false. Propositions like (F) make there be an entailment relation between what is going on in your Cartesian theatre and what is going in some region of the cosmos many miles away. There is no such entailment relation. For any

external object x, one does *not* know with incorrigible certainty that: *I am thinking that x has phi.* (One may indeed know such a truth, but not with Cartesian certainty.)

Of course, people often *say* things like "I know, beyond any doubt, that I am thinking I want to drink water", and "I know, beyond any doubt, that I want to talk to Mary". If these people are philosophers, they may even be using the words "beyond any doubt" to denote Cartesian certainty, as opposed to the probabilistic pseudo-certainty that characterizes our knowledge of the external world. These people sincerely – and, from some viewpoint, correctly -- use sentences whose literal meanings are propositions like (F). But the beliefs they are expressing through such sentences do not coincide with the literal meanings of those sentences. What they are expressing is a belief like (E). For now very familiar reasons, to *express* a belief like (E), you must use some sentence *S* that semantically *encodes* an object-involving proposition. But that object-involving proposition itself will not be a constituent of what you are expressing or of what you believe. What *is* such a constituent is the *existential* proposition through which you *grasp* the aforementioned object-involving proposition. If you *do* have incorrigible knowledge of anything, it is of the proposition that results when you give *widest-scope* to epistemic operator that you are attaching to that existential proposition.

Entailment relations and rationality: the propositional basis for our critique externalism

Inevitably, there is going to be a lot of resistance to this analysis. And, I grant, I have thus far spoken rather dogmatically. Let me give some additional, and independent, corroboration for my views.

When are P and P* different propositions? If P entails something P* does not, that is surely enough.

But, at least arguably, it is not *necessary*. After all, *1+1=2* and *triangles have three sides* entail the very same things. They are analytically equivalent. But they are different propositions.

But, I would argue, there is still *a* difference in the entailment relations associated with these propositions.

Suppose you want to prove the thesis: *some kinds of shapes have three sides.*

If your starting point is: *triangles have three sides*, it will be very easy for you to prove your thesis.

If your starting point is: *1+1=2*, you can still prove your thesis. But the proof will be much more circuitous.[111]

Your starting point *might* even be the thesis itself. In that case, the proof will be completely non-circuitous: even less so than the first of the two proofs we mentioned.

We might put it this way. There is an "analytic route" between *some kinds of shapes have three sides* and *1+1=2.* And there is also an analytic route between *some kinds of shapes have three sides* and *triangles have three sides.* But those two routes are very different.

The term "analytic route" is obviously vague. But it seems to correspond to *some* legitimate concept. Like a lawyer examining a witness, I ask the judge – you, the reader -- for some latitude. Give me this concept, and I will soon repay the loan.

With some reservations, we might say this. P_1 and P_2 are analytically equivalent propositions just in case, for any proposition P_3, the analytic route between P_1 and P_3 coincides with that between P_2 and P_3.

Let us include relations of probabilification, not just entailment, in the scope of the term "analytic route". Obviously if P_1 gives 98% probability to P_3, while P_2 gives it 12% probability, or 0%, then P_1 and P_3 are not the same proposition.

So, to a first approximation, we might say this. Given proposition P, and a proposition P*, P and P* are the same proposition exactly if, for any proposition P**, the analytic route between P and P** coincides with that between P* and P**. Let us say that, when that condition is met, P and P* have the same "analytic profile."

Is it possible for distinct propositions to have exactly the same analytic profile? It seems not. Propositions are individuated by their contents – by what they say (so to speak) and how they say it. If exactly the same things can be inferred from P that can be inferred from P*, and they are to be inferred in exactly the same way, then what difference in content is there between P and P*? If P tells you exactly the same thing as P*, and tells it to in exactly the same way, then how can P and P* be different propositions? To be a proposition just *is* to be a "telling", so to speak. It doesn’t make any sense to say that two distinct propositions tell you exactly the same thing in

exactly the same way. [112]

To be rational is to be adept at tracking analytic routes. It is to be adept at drawing inferences, deductive and inductive.

A corollary is that, if the inferences that can be rationally draw from P coincide with those that can be rationally drawn from P*, and the *way* those inferences are to be drawn in the one case coincides with the way they are to be drawn in the other case, then P and P* must be the same proposition.

Let us bring all of this to bear on externalism. Let w and w* be different "worlds". Here I am using "world" in the sense in which philosophers discussing modality use this term. So w and w* are parallel universes, not different planets.

Let us now tell a story about Smith. Until time t, Smith's life in w is exactly like Smith's life in w*. Prior to t, Smith-in-w is qualitatively identical Smith-in-w*. Further, until t, the distal causes of Smith-in-w's condition are identical with Smith-in-w*.

At t, the two worlds diverge. At that time, in w, Smith has a visual experience that is caused by Mary. At that time, in w*, Smith has a visual experience that is caused by somebody who looks and acts exactly like Mary, but isn't Mary – it is, let us suppose, Helga. (So Helga is not Mary's "counterpart" or anything of the sort. Helga is simply a different person, albeit one who is qualitatively much like Mary.)

Smith-in-w believes that the woman he is seeing is lovely. Smith-in-w* believes that the woman *he* is seeing is lovely.

Let P be the perception that Smith has at t in w. Let P* be the perception that Smith has at t in w*. So P is caused by Mary; and P* is caused by Helga.

Finally, let us stipulate that, leaving aside the distal causes of their experiences, Smith-in-w and Smith-in-w* are exactly the same before, during, and after t. They are, throughout their lives, atom-for-atom duplicates. Further, modulo the fact just described regarding Mary, the distal causes of their conditions are identical.

According to the externalist, the *content* of P is different from the content of P*. P is caused by one thing; P* is caused by a different thing. And these differences in causal origin constitute differences in *content.* The content of P has Mary as a constituent; the content of P* does not have Mary as a constituent.

Further, the externalist says that Smith-in-w and Smith-in-w* have

different beliefs. What the former believes is: *Mary is lovely*. What the latter believes is: *Helga is lovely.*

The externalist position is consistent with an important fact about what makes Smith's beliefs true. In w, what makes his beliefs true is *Mary's* being lovely. In w*, it is *Helga's* being lovely. Presumably, if some belief B has a different truth-maker from some belief B*, then B and B* are beliefs in different propositions.

Challenging the externalist picture in light of our analysis of propositions

I wish to challenge the externalist picture. First of all, is there any inference that Smith-in-w can rationally make that Smith-in-w* cannot? Is there any belief that Smith-in-w can rationally hold that Smith-in-w* cannot?

Of course, there are beliefs that Smith-in-w is *right* to have that Smith-in-w* is wrong to have. Smith-in-w, but not Smith-in-w*, would be *right* to believe: *I am having a visual experience that is caused by Mary.* So what Smith-in-w is *right* to believe doesn't coincide completely with what Smith-in-w* is right to believe.

But Smith-in-w is no more, and no less, *rational* than Smith-in-w*. In both cases, they arrived at exactly the same conclusions, in exactly the same way, on the basis of exactly the same evidence. One is wrong, one is right. But that has nothing to do with a divergence in rationality.

To be rational is specifically to be adept at seeing relations of bearing (entailment and probabilification) among the information that one has at one's disposal. (The information at one's disposal certainly include the propositions one grasps. It is a matter of controversy whether that information *coincides* with those propositions. Some hold that there is non-propositional, or "non-conceptual", information. I myself am open on this.)

It follows that, if the information at Smith-in-w's disposal is to be different from the information at Smith-in-w*'s disposal, then there must be some inference that the one can rationally make that the other cannot. But there is no such inference. So the information – and *a fortiori* the set of propositions - at Smith-in-w's disposal must coincide with that at Smith-in-w*'s disposal.

The externalist is forced to say that, even if two people grasp entirely *different* propositions, the inferences the one can rationally make

may coincide perfectly with those the other can rationally make.

But this move eviscerates our concept of rationality. Uncontroversially, to be rational is nothing more and nothing less than to be able to track relations of bearing among the information at one's disposal. Obviously the propositions one grasps are, at the very least, part of such information. To say that Smith-in-w can rationally think such and such just in case Smith-in-w* can rationally think such and such just *is* to say that the information at the disposal of the former coincides with that at the disposal of the latter.

The content of a proposition *is* its analytic profile. P and P* are the same proposition exactly if they have the same analytic profile. They have the same analytic profile exactly if, for any proposition P**, the bearing that P has on P** coincides with that which P* has on P**. (Here when I talk about "bearing", I don't just mean *whether* P entails P**: I also mean *how* it entails it.)

To sum up, in so far as Smith-in-w and Smith-in-w* are permitted to make exactly the same inferences, it follows, truistically, that they grasp exactly the same propositions. With one qualification – a qualification that in no way helps externalism -- they are permitted to make exactly the same inferences. Therefore (modulo that one qualification: see next paragraph) they grasp exactly the same propositions. If we say otherwise, we strip the concepts *proposition* and *rationality* of any content.

There is a subtlety. Uncontroversially, Smith-in-w and Smith-in-w* don't grasp *quite* the same propositions. Suppose Smith-in-w thinks to himself *I am tired,* and that Smith-in-w* also thinks *I am tired.* Smith-in-w's thought is true exactly if Smith-in-w is tired. Smith-in-w*'s thought is true exactly if Smith-in-w* is tired. So they have had different thoughts.[113] So, to that extent, what Smith-in-w thinks is going to differ from what Smith-in-w* thinks, no matter how perfect the *qualitative* similarity between them is.

But notice that this has nothing to do with externalism. If Smith-in-w thinks *I am tired*, his thinking that supervenes entirely on his intrinsic properties. The same is true (mutatis mutandis) of Smith-in-w*. You can "spin the possible worlds"[114] as much as you please. Whether a person is thinking *I am tired* is entirely a function of his intrinsic properties: a brains in vats, and dupes of Cartesian demons, are just as capable of thinking that as their embodied counterparts. Those thoughts are externalism-proof. So the just mentioned divergence between Smith-in-w and Smith-in-w* has nothing

to do with their respective environments having a constitutive role in what they think.

But with that one qualification, Smith-in-w and Smith-in-w* are permitted exactly the same inferences. So exactly the same information, and (a fortiori) exactly the same propositions, are at their respective disposals. Thus the differences in their respective environments is irrelevant to the contents of those thoughts.

Truth-maker versus content

Still, those environmental differences *do* have great – an *incalculable* _-- importance. One's environment is what makes one's thoughts be *true* or *false*.[115] It provides truth-makers and falsifiers. But, *pace* externalism, one's environment doesn't to any degree fix the *content* of one's thoughts. Recall what we said earlier:

(s) "Sally met a man"

is made *true* by the fact that Sally met Frank. But (s) is not about Frank, and neither is the proposition encoded in (s). Suppose I have a hallucination of an elephant. My visual experience tells me something false; it says to me: *there is an elephant with such and such properties in thus and such a place.*…Suppose my doppelganger has a qualitatively identical visual experience, but his experience is caused (in the right way) by Abby the elephant. It would be madness to say that, under these circumstances, my visual experience bears no information in common with my doppelganger's. Obviously they *do* bear common information. Mine tells me: *there is an elephant over there.*…And my doppelganger's visual experience tells him exactly that. But my doppelganger's experience is *true* (veridical). It is made true by the fact that Abby is in the vicinity. My experience is false.

There is no "blank" in my visual experience; it is not as though, in my visual experience, there is some elephant-shaped cavity. No – my visual experience is quite complete. That is why it is false; that is why people can tell me I'm seeing things *wrongly*. Things with gaps in them – propositional-functions – aren't true or false. Propositions have that privilege. My visual

experience, being false (and therefore either true or false), is (in that respect, though perhaps not in others[116]) proposition-like, and not propositional-function-like. The problem with my visual experience is not that it has a gap in it. The problem is that what it says is false.

My doppelganger's visual experience is no more, and no less, *complete* than mine. There are no empty spaces in mine that are filled in his. Mine is wrong; his is right; they are both complete, both "saturated". If his perception were *more* complete than mine, in virtue of being caused (in the right way) by an elephant, that would imply that mine was *less* than complete, that there were blanks in mine (free-variables, so to speak). But, as we've just seen, mine is quite complete: there is no elephant-shaped cavity in mine. Therefore there is, so to speak, no *filled* elephant-shaped cavity in his.

A veridical perception is not something that is like a hallucination except that the blanks have been filled in. Given a veridical perception P, and a hallucination H that is qualitatively just like P, both P and H are equally "saturated". So P is not what results when something has been added to the *content* of H. P is what results, so to speak, when H is *made true*, not when H's content has been supplemented.

If I am not mistaken, one of the two root-problems with externalism is a failure to distinguish the content of a thought from its truth-makers. There is another root-problem with externalism. Let us now discuss this other problem.

Another argument against content-externalism

I believe that, ultimately, there is but one confusion underlying content-externalism: a failure to distinguish the information encoded in our perception from the information which we derive by making *judgments* about our perceptions. What I *see* is one thing. The judgments that I make about what I see are another thing. I don't *see* quarks. But on the basis of what I see, I make judgments to the effect that there are such things.

We must distinguish perceptual content from *meta*-perceptual content. Externalism is what results when these very different things are collapsed.

Given *only* the information encoded in my perception, I have no way of knowing whether the cause of my perception is water or twin-water or some third substance. In light of other perceptions of mine, some judgment

on that matter may be warranted; those other perceptions may warrant the *judgment* that the cause of my perception and this as opposed to that chemical structure. But if we focus on what that the perception itself tells me, I have no reason to believe anything about its chemical structure – not even that it has a chemical structure.

Suppose that you have a perception of Chip. Nothing *in* the perception tells you that it is Chip, as opposed to twin-Chip. Given only the information encoded in the perception, it would be irrational of you to conclude that you were seeing Chip as opposed to his twin, or *vice versa.*

Later perceptions may warrant the conclusion that Chip, and not twin-Chip, was the cause of that perception. But nothing in the perception itself warrants that conclusion. Indeed, if you were to come to that conclusion solely on the basis of the information encoded in that one perception, you would be guilty of deep irrationality.

What decides the matter – what makes it rational to conclude that Chip, as opposed to twin-Chip, was the cause of your perception – lies in *other* perceptions that you had. So even though the perception was *in fact* caused by Chip, and not twin-Chip, that fact is not reflected in the information[117] borne by the perception itself. So the information borne by Chip perception of Chip is identical with that borne by perceptions of twin-chip. Thus, Chip is no more a constituent of the information borne by Chip-perceptions than twin-Chip is. Obviously twin-Chip is not such a constituent; therefore neither is Chip. Chip is a *cause* of that perception; and later judgments may warrant your believing that Chip, and not twin-Chip, has that honor. But the point is that nothing in the perception itself – nothing in what that perception, considered by itself, tells you – warrants the view that Chip, as opposed to twin-Chip or some third entity, is what caused it. What warrants that view is not the perception itself, but some *judgment* made about it, in light of the information borne by other perceptions.

What makes it rational for a scientist to conclude that some substance has this or that chemical composition is obviously not the information encoded in some one perception: *that* information is neutral on the question whether the substance is xyz or ABC. So even though the perception was *in fact* caused by xyz, and not ABC, that fact is not reflected in the information borne by the perception itself. This means that

The externalist says:

(ex) Chip himself is a constituent of the representational content of your perception.[118] *The information encoded in your perception includes Chip, and not – at least not merely – information that applies to Chip. Water itself is a constituent of the representational content of your perception; that information doesn't merely apply to water.*

But if (ex) were true, then perceptions of Chip would contain different information from perceptions of twin-Chip: the one batch of information would have Chip as a constituent, while the other batch would not. If *that* were the case, then what a perception of Chip tells you would be different from what a perception of twin-Chip would tell you. But given a perception of Chip – given only the information encoded in that perception – you have no way of knowing whether it was caused by Chip or twin-Chip: there is *nothing* that, taken by itself, a Chip-perception tells you that a perception of twin-Chip does not tell you.

Why externalism cannot accommodate the causal potency of the mental

I myself think that all the arguments for content-externalism are completely neutralized by making a few key distinctions: literal versus cognitive meaning, perceptual versus meta-perceptual information, content versus truth-maker.

Be that as it may, practically everyone these days accepts some form of content-externalism. But, leaving aside everything we've so far in criticism of it, there is an obvious problem with content-externalism: it makes the representational contents of our thoughts be causally inert -- completely and totally without the ability to *do* anything. If externalism is right, the fact that I am thinking *the number two is an even prime* has no causal force; it has nothing to do with the fact that, subsequently, I think *there is at least one even prime*. The only properties of my organism that have any causal powers are the non-representational ones.

By definition, content-externalism says that what you are thinking does not supervene on the intrinsic properties of your organism. If we leave aside facts about the *causes* of their conditions, X and Y can be in exactly the same condition and be having utterly different thoughts. X is thinking about

Mary. Y is thinking about twin-Mary. X is thinking about Alpha Centauri. Y is thinking about *nothing*: in his universe, there is Alpha Centauri – so his would-be Alpha-Centauri thoughts are blanks: they contain a star-shaped blank, where X's thoughts contain a star.

The causal properties of one's brainstates obviously supervene on their local properties. If two people have qualitatively identical brains, their brains will have exactly the same causal properties (modulo any indeterminism relating to quantum mechanics). It is totally irrelevant whether the one person's brain bears such and such relation to Socrates, while the other person's brain does not bear that relation to Socrates.

This must be understood aright. Obviously one's brain is causally responsive to objects that are not in its immediate vicinity. Everything affects everything. The moon affects the tides. The tides affect the climate. Because I am swimming or am a farmer who cares about the weather, the climate affects my brain-state and thus my mental state. The content externalist is not uttering the platitude that things are never totally causally isolated from one another. He is saying: *stipulate* that, leaving aside facts about the origins of their conditions, X and Y have qualitatively identical brains and bodies. Also suppose that X's condition has such and such etiology and Y's condition has thus and such etiology. So, by stipulation, the etiologies of their conditions differ (in some specified way) and, also by stipulation, those conditions (leaving aside facts about their origins) are qualitatively identical. In that case, the content-externalist says, X and Y are having different thoughts: one is thinking *Alpha Centaury is lovely*; the other isn't having a complete thought at all; he is having, at most, the pseudo-thought *__is lovely*. So the content-externalist is saying: even if we leave aside the fact that no two things are causally isolated – even if stipulate that X's condition is identical with Y's condition and also stipulate that the etiology of X's condition differs in thus and such way from the etiology of Y's condition – *still* what X is thinking is simply different from what Y is thinking.

Externalism practically amounts to the claim the causally effective properties of a person are the ones that have no representational content. Content supervenes *not* on the local, causally effective properties of one's organism, but on facts about the causally irrelevant, spatially and, often, temporally distant facts about the cosmos. So suppose you are thinking P: if content externalism is right, the fact that you are thinking P is irrelevant to the subsequent course of your brain-states and, therefore, of your thoughts.

Of course, some self-described externalists concede that, in at least some cases, representational content, of a certain kind, has certain causal powers. But, I will argue, in so far as one makes such a concession, one isn't really an externalist at all.

Suppose I think *two is an even prime.* Let B be the brain-state mediating this thought. B will almost certainly lead to my thinking *there is at least one even prime*. Let us suppose this is just what happens. Let B* be the brain-state mediating this second thought.

What is going on here is that my thinking one thing leads to my thinking another thing that has a related content. The fact that B has a certain representational content is causally effective. To a very high degree, our thought-processes are consistent with the representational contents of those thoughts. I think *something is tall* because I think *Socrates is tall.* The fact that I was having a thought with a certain content leads me to have a thought with another, related content.

This isn't to say that the neural and chemical properties of brain are irrelevant. They obviously are relevant - as relevant as can be. It is to say that the representational properties of brain-states are realized by, or coded in, their causally effective physical (neural, chemical, atomic) properties. The causally effective properties of brain-states are their local ones: facts about the atomic, molecular, cellular structure, their electrical properties, and so on. What is *not* effective is anything buried in remote regions of time or space. The representational properties of brain-states are causally effective. Therefore, the representational properties of brain-states must supervene on, or be identical with, their aforementioned local (non-distal) properties. Externalism must be wrong; it makes representational content have no causal role or, at least, have an overly attenuated causal role.

What we *do* think is consistent, to an extremely high degree, with what we *should* think, given the representational contents of our thoughts. The obvious way to explain this is to say that the representational properties of our thoughts are causally effective. Others have proposed different explanations, as we will presently see. But this is surely the one we want to hold onto; and, arguably, it is the only coherent one.

For the sake of a *reductio*, suppose that the representational contents of our thoughts are without effect. In that case, *ex hypothesi*, our *grounds* for believing something never have anything to do with our *actually* believing it. So *ex hypothesi* any one of our beliefs – whether true or false – is in some

sense groundless. There might *be* grounds for it; but our *believing* it ends up being groundless.

Let me elucidate this distinction. Suppose Bill has two fruits (one apple and one orange) I form the belief that Bill has two fruits. No doubt, there are objective grounds for believing this about Bill. But if the representational contents of my mental states are causally inert, then *my* believing it has nothing to do with those grounds. So *if* the representational content of our mental states are causally inert, then all of our beliefs are groundless, including (if this happens to be among our beliefs) the belief that the representational contents so our mental states are causally inert. So the view that such contents are inert, if true, cannot rationally be held. (This doesn't mean that it's false; but it means that, by its own lights, it cannot be rationally held.)

In a word, content-externalism seems to make representational content have no causal powers. This is practically the same as making *mental entities* have no causal powers. For representationality is, at the very least, an extremely pervasive feature of the mental.

A distinction must be made. Even if we identify mental entities with brain-states, and even if we strip mental entities of causal powers, it doesn't follow necessarily that we have stripped those brain-states of causal powers. What has causal properties is never, strictly speaking, an object: it is a state of affairs involving an object. What shatters the window is not the rock: it is a state of affairs involving the rock; it is the rock's having a certain mass, a certain structure, along with its moving with a certain velocity. Similarly, by itself, no brain-state causes anything. What causes something is a state of affairs involving a brain-state: a brain-state's having certain morphological, chemical, atomic, or representational properties. If we strip mental entities of causal powers, we are not saying that the brain-states which mediate thoughts are causally inert in *every* respect; we are still allowing (what is undeniable) that a brain-state's having such and such chemical properties is causally effective; what we are denying is that a brain-state's having such and such *representational* properties is causally effective.

But for various reasons – ranging from the purely instinctual to the utterly logical – people want to hold onto the idea that what we think has causal powers. (I want to hold onto that idea.) Externalism seems deeply inconsistent with this idea.

Defenders of externalism have reacted in a few different ways to this

problem. The main reaction has been to have no reaction – to ignore the problem – and to pursue different issues, untroubled by the rhinoceros in the room.

Another reaction – one that appears very reasonable -- is this. There are two kinds of representational content: "narrow" and "wide". Narrow-content supervenes on the local, causally effective properties of brains. Wide-content does not – wide content supervenes on the local properties *plus* the potentially vast swath of space-time comprising the relevant facts about the origins of those local properties.[119]

This move, I would argue, is correct *only if* by "narrow content" one means *content*, and by "wide-content" one means *truth-maker*, as opposed to content proper. In other words, this move works only if is taken as a bit of *anti*-externalism. If this move is taken in any other way, then it ends up being simply false, as it then wrongly turns the kind of content that does supervene on local facts about one's organism into something isn't true or false at all.

Let Bob be your atom-for-atom duplicate. But suppose that Bob's waking life is one giant hallucination. Not because he is mentally defective – he is as sane as our best accountants, and the mechanisms in his brain that are responsible for sensory experience are perfectly intact. Bob's problem is that, because of how he is embedded in his environment, all his sensory experiences come about in the wrong way; they end up being mis-perceptions or hallucinations. Through no fault of his own, Bob "sees" an elephant where there is no elephant.

We've already seen that Bob's experiences *do tell him something.* Let Tim be a victim of Descartes' evil demon. Obviously Tim's experiences tell him something. What they tell him is false. (They tell him he is seeing a hand, when he is not seeing a hand.) And even if, by coincidence, what they tell him is true, it isn't knowledge, since it came about in the wrong way. But his experiences still tell him something. What Tim's perceptions tell him is *false* or only *coincidentally* correct; it isn't neither true nor false. The same is true of Bob. So "narrow-content" – the kind of content that survives the wrath of a Cartesian demon – is true or false. Exactly the same thing is true of Bob.

As we've seen, given something that is true or false, there is a sense in which nothing can be *added* to it; there is a sense in which such a thing is entirely complete, and contains no free-variables, no open-parameters. Of course, things can be added *on*to it. If you think *two is even*, you can embed that thought in other thoughts (*two is even and snow is white, therefore there*

is something x such that snow is white and x is even …) But nothing can be added *in*to it. Nothing true or false contains free-variables, or any analogue thereof.

This scuttles the idea (no longer widely held) that internal content corresponds to Kaplan's "character".[120] Narrow content is true or false; the character of a sentence-type is not true or false; it is a *function* from contexts to things that are true or false, but it is not itself true or false.

One could take the bullet-biting strategy of saying:

Bob and the victim of Descartes' demon don't have sensory experiences that are either true or false. They don't even make it that far. What Bob's waking experience tells him isn't true or false; it is so fraught with gaps that it doesn't tell him anything true or false. What it tells him has free-variables in it, and has the form: __looks tired and very much not up for any sort of confrontation; but__is now approaching__ and, judging by the expression on__'s face__isn't going to be very happy about__'s little soiree with__'s wife, even though it happened with__'s consent…[121]

But one's sensory experiences -- even if hallucinatory, even if lacking the right kind of causal connection to the outside world – are obviously replete with existential content. Surely the victim of a hallucination is being told something *false* (or, at best, coincidentally correct); he isn't being told *nothing*. After all, the hallucinator is misled; he is told wrong (or, at best, coincidentally correct) things by his experiences; so he is not being "told" something that has free-variables in it. Narrow-content, supposing that there is such a thing – supposing that there is any distinction between it and any other kind of content – contains no free-variables or anything analogous.

In light of this, what becomes of *wide*-content? Wide-content, if there is such a thing, is not the result of "filling out" narrow-content, of binding (or substituting for) the free variables in the latter; for there are no such variables. Wide-content would have to be an *extra* kind of content. So my perceptions, and my thoughts, would have *two* different contents, both truth-evaluable.

So then we are saying that our perceptions, and thoughts, are *ambiguous*. But that isn't quite right. The connection between the two kinds of content is tighter than that. The narrow kind consists in existential information. The wide kind consists of what we might call *singularizations* of

such information. (A singularization of *Sally met a man* would be *Sally met Harold*.) Idealizing away from issues relating to the question whether sense-perceptions encode propositions (or otherwise digital information), the narrow content of my perception, and the belief I have in virtue of having that perception, is something like: *there is a creature in front of me with giant fangs that is growling at me…*

The *wide*-content of my thought and belief is a singular proposition which *makes true* that existence claim; it is something like: *Fido is growling at me.* So "wide-content" becomes the truth-maker of "narrow-content".

But if we take this route, we are really abandoning externalism; we are simply re-stating, under the cloak of new terminology, the point that the *truth-maker* of a thought is one thing, and its *content* is another. Remember what we've said time and time again: the *content* of the proposition *Sally met a man* has nothing to do with Fred, even though Sally's meeting Fred is what makes that proposition be true. That proposition is true in worlds where there is no Fred, where Sally meets Bob instead. The content of an existence claim is one thing; its truth-maker – which, ultimately, is always a singularization of that proposition -- is another thing entirely.

The wide-content/narrow-content ploy is thus available only to the *anti*-externalist. If you concede that thoughts have narrow content, you must concede, for reasons earlier given, that those narrow contents are true or false and are also existential. If you then say that our thoughts have "wide-content", you must say that the wide-contents are really the *truth*-makers of those existence claims. But if you say that, you are really saying that "wide-content" is not *content* at all.

Reconciling the truth semantic-externalism with the falsity of content-externalism

But we can have our cake and eat it too. Kripke and Kaplan and other direct reference theorists are exactly right that the contents of our *sentences* do not supervene on what is "in our heads". So *semantic* externalism is quite right. Putnam is quite right that what makes Oscar's thoughts *true* is that H_2O is wet; and that what makes twin-Oscar's thoughts *true* is that ABC is wet, notwithstanding that (aside from the facts about the origins of their conditions) Oscar and twin-Oscar are qualitatively identical. That is why the of success empirical theories – theories about the

microstructure of water, light, and so on -- turns on things that simply don't supervene on our intrinsic conditions.

Suppose Oscar thinks: *there is some x such that I bathe in x, drink x, and swim in x, and my theory is that x is composed o f* H_2O. Oscar's theory is correct. Suppose twin-Oscar thinks: *there is some x such that I bathe in x, drink x, and swim in x, and my theory is that x is composed o f* H_2O. Twin-Oscar's theory is wrong; for, in twin-Oscar's case, x has chemical structure ABC. The content of Oscar's thought is the same as the content of twin-Oscar's thought. But the truth-makers are different. And that is why twin-Oscar's thought, his theory, is wrong. What makes a thought *true* isn't necessarily identical with the content of a thought (the two pull apart when the thought is existential, for an existential thought is always made true by a *singular* fact); and what we think isn't necessarily what our sentences mean (this is the case when the sentences encode singular propositions, for we grasp such propositions through existential/non-singular ones). When we allow for these facts, the facts which motivate externalism can be accommodated without adopting the desperate view that the representational contents of our brain-states, narrowly individuated, are either causally inert or are not truth-evaluable.

Disjunctivism and the individuation-conditions for thoughts

Let twin-you be a brain in a vat that is qualitatively just like your brain. When you actually see an elephant, twin-you has a hallucination qualitatively identical with your veridical perception. According to John MacDowell, the representational contents of your states have *no representational content in common with your twin's states.* The evidence that is available to you is, by hypothesis, utterly indistinguishable from the evidence that is available to twin-you. For all you know, *you* could be twin-you. Like every philosopher since Descartes, MacDowell is aware of this. But MacDowell boldly says that, even so, what your experiences tell you – their representational contents – is utterly *disjoint* from what your twin's experiences tell him (which is, strictly speaking, nothing: in terms of representational content, his waking life is one big blank even though, in terms of phenomenological content, it is as rich as yours). For this reason, MacDowell's position is known as "disjunctivism".

From some point of view, externalists are right to be disjunctivists.

As we've seen, if we allow that you and twin-you share any content, then we end up saying that they share *all* content (modulo the point about…*I*…-- thoughts made earlier); we end up saying that their thoughts differ not in respect of content, but in respect of what their *truth-makers* are.

A corollary of disjunctivism is that representational content is completely inert. Uncontroversially, what is causally effective is what is common to you and twin-you. Disjunctivism says that you and twin-you share no representational content. So disjunctivism entails that representational content is inert.

Another failed compromise

There is another possible response to the dilemma for externalism that we've been discussing. One could say that where *some* thoughts are concerned, representational content supervenes on what is local, but where other thoughts are concerned, this is not the case. Consider the proposition *two is an even prime.* It is very hard to believe that twin-you doesn't grasp this proposition. Surely when we are dealing with purely conceptual truths – e.g. *triangles have three sides, the existence of money presupposes the existence of animate beings, x and y are causally related only if one of them precedes the other, there are infinitely many primes* -- it makes dubious sense to say that content supervenes on what obtains in remote regions of the cosmos.

Suppose you die and go heaven, where God tells you that, throughout the entirety of your life, you were the victim of Descartes' demon. You would conclude that there had never been a cousin Fred or an aunt Ethel. But you would surely *not* conclude that you had never known that triangles have three sides or that the existence of money presupposes the existence of animate beings.

So, quite uncontroversially, externalism doesn't apply (and isn't always intended to apply) to purely conceptual thoughts. If it applies to anything, externalism applies to mental states that represent external objects: chairs, rocks, distant stars. Given this, it might seem reasonable for a (sometime) externalist to say this:

(*) Where purely conceptual thoughts are concerned,

representational content supervenes on what is local, and is thus causally effective. Where thoughts about the external world are concerned, representational content doesn't supervene on what is local, and isn't causally effective. Whatever causal powers your Alpha Centauri thoughts have, their having those powers has nothing to do with what they represent.

We've already seen one reason to reject (*): obviously the representational content of *any* thought is causally effective; it doesn't matter whether its verifier is a fact about a distant star or a fact about the number two.

But leaving that aside, (*) seems curiously arbitrary. Given a purely conceptual thought, it is obviously an *essential* fact about it that it has a certain representational content, and also that, in virtue of having that content, it has certain causal powers. Let t be your thought that two is an even prime. There is no epistemically possible world where t has a different representational content; and there is no epistemically possible world where t doesn't incline its owner to believe that *something is an even prime* and *two is even* and *a number less than three is even*; there is no epistemically possible world where that very thought's causal properties don't, at least to some degree, track its representational content. These facts are constitutive of that thought's identity. If (*) is right, then matters are completely different where other thoughts are concerned. Their representational contents are totally divorced from their causal properties; causal and representational properties vary freely with respect to each other. Are we to say that what is essential to *some* thoughts is inessential to others? Thoughts are thoughts. What is essential to some must be essential to all.

Let t* be some thought you are having. Let B* be the brain-state that realizes t*. Assuming that some kind of materialism is correct, there is no possible world where B* exists where t* does not. But there are worlds where B* exists where Alpha Centauri doesn't exist, where Mary doesn't exist, where Robert DeNiro doesn't exist. Given any external object O, there are possible worlds where B* exists and O does not. So there is some world where B* exists and Alpha Centauri does not. After all, the individuation-conditions for brain-states don't have anything to do with Alpha Centauri. In any possible world where the facts right here on Earth are in order, there will be a B*: as for Alpha Centauri, we can take it or leave it. There is obviously

no essential, no in-all-possible-worlds, connection between B* and Alpha Centauri (except in so far as there might be some such connection between any two objects, given the high degree of causal interconnectedness of things). So if (*) is correct, then (for at least some admissible values of B*), B*'s representational content includes Alpha Centauri itself (as opposed to some concept that uniquely applies to Alpha Centauri). Recall that, in any world where there is B*, there is also t*. It follows that there is some world where B*, and therefore t*, has a totally different representational content from the one it has here and worlds where it has *no* representational content (where it isn't true or false). This means that B*, and therefore t*, does not have its representational content *essentially*; it has it quite contingently. But to be a thought just *is* to be something that has representational content. So I don't see how thoughts could fail to have their contents essentially. t* is a thought exactly to the extent that it is something that has representational content. So when we say that B* could have had some other representational content, or no (truth-evaluable) representational content, we are saying: B* could have been a numerically different thought. It must be clear what this amounts to. The externalist is committed to saying that B* could have had a different representational content. Thoughts just *are* things that bear such content. So thoughts are individuated by their representational contents. So in so far as B* could have *had* a different representational content, it would have *been* a different entity. This amounts to saying: B* might not have been B*; in some world, B* -- that very entity – is something other than B*. But this, as we know from Kripke (and others) is absurd. The only way out of this, for the externalist, is to say that thoughts are *not* individuated by their representational contents. But, again, I don't see what there is to being a thought other than having a representational content; and this makes it hard to see how the connection between a thought and its representational content could be contingent.

The arbitrariness of the externalist's position

As we remarked earlier, *some* thoughts clearly have their representational contents essentially; and it seems extremely arbitrary to say that some thoughts have this property, while others do not. It is like saying:

(#) Some chemical substances have their molecular structures

essentially, while others do not; some individuals have their conditions of origin essentially, while others do not.

I wish to distance myself as much as possible from (#), and also from its externalist counterpart.

Chapter 20 Literal versus communicated meaning

We must zealously distinguish what is cognitively *grasped* from what it is *semantically encoded* in our words. I think that a failure to make this distinction lies at the bottom of all the conundrums associated with externalism.

Recall our friends Smith-in-w and Smith-in-w*. Remember, up until time t, everything about them – their conditions, and also the distal causes of their conditions -- is the same; and *after* time t, leaving aside the distal causes of their conditions, everything about Smith-in-w and Smith-in-w* is the same. The only difference between them is this. At t, Smith-in-w has a visual experience that is caused by Mary. At that same time, Smith-in-w* has a visual experience that is caused by Helga.

Let us continue our story. At t, Smith-in-w points to the woman he is seeing and says to a companion: "let us refer to that person as ‘Roseanne’". At t, Smith-in-w* points to the woman *he* is seeing and says (to a companion): "let us refer to that person as ‘Roseanne".

In w, "Roseanne" refers to Mary. In w*, "Mary" refers to Helga. So, in w, the literal meaning of

(i) "Roseanne is lovely"

is:

(ii) *Mary is lovely.*

But in w*, the literal meaning of (i) is:

(iii) *Helga is lovely.*

(ii) and (iii) are, unquestionably, entirely distinct propositions. And, unquestionably, what is *literally* meant by an utterance of (i) in w is entirely distinct from what is literally meant by such an utterance in w*.

As we saw earlier, when you point to some object O and say: "that thing is named ‘Smith’", you are really saying:

Sentences of the form

"…Smith…"

mean:

…*O*….

So, in w, (i) means (ii), whereas in w*, (i) means (iii).

What is literally meant by *sentences* does not supervene on what is "in people’s heads": facts about the distal causes of one’s condition *are* individuative of *literal meaning.*

But for reasons we have seen all too clearly, the information that an utterance of (i) will *communicate* to Smith-in-w will coincide with the information that such an utterance will *communicate* to Smith-in-w*. In both cases, the very same existence claim will be the thing *communicated.*

Here, no doubt, the following will be said:

You are overlooking the most important thing of all. Consider the proposition that Smith-in-w associates with utterances of (i). Let @ be that proposition. @ will be made true by Mary’s being lovely. It is irrelevant if Helga or Susan, or anyone other than Mary, is lovely.

Now consider the proposition that Smith-in-w associates with utterances of (i). Let @* be that proposition. @* will be made true by Helga’s being lovely. It is irrelevant if Mary or Susan, or anyone other than*

Helga, is lovely.

So what makes @ be true is something different from what makes @ be true. Propositions are individuated by their truth-conditions. Therefore @ and @* are different propositions. So you are wrong. Externalism is correct after all.*

Remember the point we made earlier. There is a difference between a proposition's being *about* something and its being *made true* by some fact involving that thing. Let p be the proposition encoded in "Sally met a man". p is *made true* by the fact that Sally met Frank, or Dave, or Harry. But p is not about any of those men. A given existence claim can be *made true* by different things. In some worlds, p is made true by the fact that Sally met Frank. In others, it is made true by the fact that Sally met Dave. Obviously p doesn't vary from world to world. And (less obviously) the truth-*conditions* of p don't vary from world to world. The truth-*maker* of p varies from world to world. In every world, the truth-*condition* for p is that there be *some* man x or other such that Sally meet x. This condition can be satisfied by Sally's meeting Dave or by her meeting Frank. So the thing that *satisfies* the condition can vary. (Obviously different things can *satisfy* the condition: *x is human*.) But the condition itself – the truth-condition associated with the proposition – is world-invariant.

Let us talk again about **@** and **@***. For reasons we've seen, **@** is some existence claim. And, for reasons we've seen, **@*** is the very same existence claim. In w, **@** is *made true* by the fact that Roseanne is lovely. In w*, **@** (i.e. **@***) is *made true* by the fact that Helga is lovely. But in both cases, it is **@** that is grasped and that is made-true. The truth-*maker*, not the truth-*condition*, is what varies from world to world.

Let us relate these points to Putnam's[122] deservedly celebrated thought-experiment. Suppose that Jones lives here in w. There is a certain liquid that Jones drinks, bathes in, etc. Jones and his fellows refer to it as "water." Water is H_2O.

Let w* be some other world. Jones-in-w is a molecule for molecular duplicate of Jones-in-w*. So, of course, in w* there is a liquid that Jones bathes in, drinks etc. But in w*, that liquid is xyz, not H_2O. Remember that, apart from that one fact, everything in w* is just as it is here. So, in w*, Jones and his fellows refer to that liquid as "water".

To simplify things, let us suppose that Jones (in both worlds) is living well before anyone knew anything about the chemical compositions of substances.

When Jones-in-w says "water quenches thirst", the proposition that is *literally meant* by his words is true exactly if H_2O quenches thirst.

When Jones-in-w* says "water quenches thirst", the proposition that is *literally meant* by his words is true exactly if *xyz* quenches thirst.

There is no doubt that, in w, "water" refers to H_2O; no doubt that, in w*, "water" refers to xyz.

Suppose you point to some liquid and say: "let us refer to that liquid as ‘water’". What you are saying is, in effect:

(*) There is some substance *s* that I am indicating. Whatsoever *s* should turn out to be, "water" refers to *s*.

Because the descriptive information in (*) is given wide-scope with respect to the "refers to" operator, "water" is a rigid designator – indeed, a *de jure* rigid designator. It is not an indexical[123]; it is not a definite description.[124] It no more varies in reference than "Richard Nixon".[125]

So there is no doubt that the proposition literally meant by

(iv) "water quenches thirst"

in w is different from the proposition literally meant by (iv) in w*.

But literal meaning is not always communicated meaning. And, for reasons we’ve discussed, the information which Jones-in-w associates with (iv) is identical with that which Jones-in-w* associates with it.

Let *i* be the information that Jones-in-w associates with (iv), and let *i** which Jones-in-w* associates with it. On the basis of *i*, is there any inference which Jones-in-w* can rationally make that Jones-in-w* cannot? No. *i* warrants any conclusion that *i** warrants, and *vice versa.* So *i* and *i** are the same piece of information. The proposition that (iv) *communicates* to Jones-in-w is identical with the one it communicates to Jones-in-w*. Of course, in w, (iv) semantically encodes one proposition in w and a different one in w*. But that fact, as we have seen, is irrelevant.

Now there is *an* important difference between Jones-in-w and Jones-in-w*. In w, the thought that Jones associates with (iv) will be *made true* by

the fact that H_2O (water) quenches thirst. In w*, the thought that Jones associates with (iv) will be *made true* by the fact that *xyz* (non-water) quenches thirst. So the *truth*-makers of Jones-in-w's thought will be different from the truth-maker of Jones-in-w*'s thought. But the truth-*conditions* are the same. In both cases, the thought in question is some existence claim; in w, it is made true by water; in w*, it is made true by xyz. But what the thought is about is world-invariant – just as what the proposition *Sally met a man* is about is world-invariant, even though it has different truth-makers in different worlds.

The divisive passages from Kripke

Kripke made a powerful (in my view, a conclusive) case that "Hesperus", "Phosphorous", and other names are not descriptions and are mere labels.[126]

Kripke also "discovered" (if that is the right word) that some truths are necessary *a posteriori.* "Hesperus is Phosphorous" is necessary: after all nothing could have been anything other than itself. But it is a posteriori: we had to use telescopes to learn it; pure reason wasn't enough.

These discoveries are in tension with each other. If "Hesperus" and "Phosphorous" really are labels, and they label the same thing, then "Hesperus is Phosphorous" should encode the same proposition as "Hesperus is Hesperus". We've already dealt with the possibility that those sentences are synonymous, and with the *apparent* problems with that view.

There is an even more interesting internal tension within *Naming and Necessity.* In some places, it seems as though Kripke is *denying* the existence of necessary a posteriori truth. I will quote two of the relevant passages:

The objector is correct when he argues that if I hold that this table could not have been made of ice, then I must also hold that it could not have turned out to be made of idea; it could have turned out that P entails that P could have been the case. What, then, does the intuition that the table might have turned out to have been made of ice or of anything else, that it might even have turned out not to be made of molecules, amount to? I think it means simply that there might have been a table looking and feeling just like this one and placed in this very position in the room, which was in fact made

of ice. In other words, I (or some conscious being) could have been qualitatively in the same epistemic situation that in fact obtains, I could have the same sensory experience that I in fact have, about a table which was made of ice…

The general answer to the objector can be stated, then, as follows: Any necessary truth, whether a priori or a posteriori, could not have turned out otherwise. In the case of some necessary a posteriori truths, however, we can say that under appropriate qualitatively identical evidential situations, an appropriate corresponding qualitative statement might have been false. The loose and inaccurate statement that gold might have turned out to be a compound should be replaced (roughly) by the statement that it is logically possible that there should have been a compound with all the properties originally known to hold of gold. The inaccurate statement that Hesperus might have turned out not to be Phosphorous should be replaced by the true contingency mentioned earlier in these lectures: two distinct bodies might have occupied, In the morning and the evening sky, respectively, the very positions occupied by Hesperus-Phosphorous-Venus.[127]

In these passages, what Kripke seems to be saying is this:

When we say things "Hesperus might not have been Phosphorous", we aren't really saying that Hesperus might not have been Phosphorous; we are making a claim about our epistemic situation; we are talking about the information through which astronomical information is given to us. How do we know about the planets and the stars? Through various bits of evidence involving telescopes, photographic plates, and the naked eye. Ultimately we know about the heavens through sense-perceptions. Consider those perceptions that relate, in a certain way, to the presence of a last body to remain in the morning sky. Let us, loosely speaking, refer to them as your "Hesperus-perceptions". Now consider those perceptions of yours that relate, in a certain way, to the presence of a first celestial body to appear in the evening sky. We might, loosely speaking, refer to those as your "Phosphorous-perceptions". There is some world where something X creates perceptions just like your Phosphorous-perceptions, and where something Y, not identical with X, creates perceptions just like your Phosphorous perceptions. So there are worlds in which your evidential situation is perfectly reproduced, but in which that evidential situation, if traced back to

its roots, turns out to correspond to the existence of two distinct planets. When you say "Hesperus might not have been Phosphorous", you aren't saying Hesperus might not have been Phosphorous: such a claim is absurd. A thing is what it is, and isn't something else. So a claim of the form "x could have been something other than y" amounts to a triviality (if x and y are different) or an absurdity (if x and y are identical).

So "Hesperus might not have been Phosphorous" *is*, strictly speaking, absurd or trivial. Therefore, "Hesperus is Phosphorous" is, strictly speaking, trivial or absurd (depending on whether it is true or false). What is *not* trivial or absurd is not some claim about planets; what is not trivial or absurd, and is in fact true, is some proposition concerning the origins of various evidences we have, some proposition that is (very approximately) like:

(*) Let $H_1 \ldots H_n$ be those perceptions of yours that relate, in a certain way, to the presence of a unique celestial body that disappears after all others from the morning sky. Let $P_1 \ldots P_n$ be those perceptions of yours that relate, in a comparable way, to the presence of a unique celestial body to appear before all others in the evening sky. There are possible worlds where something causes you to have perceptions qualitatively just like $H_1 \ldots H_n$, and where something Y, distinct from X, causes you to have perceptions qualitatively just like $P_1 \ldots P_n$.

So the proposition:

Hesperus might have been something other than Phosphorous

is trivial or absurd. Therefore, the same is true of *Hesperus is Phosphorous.* What is *not* trivial or absurd is (*) – is some proposition concerning our epistemic situation.

(*) is not "necessary a posteriori". (*) is a form of the old skeptical claim: given any piece of evidence E that you have regarding the external world, it is logically possible that E should have come about in some way other than how it did come about. Let's say that E is a perception of yours – a perception of the kind that, if veridical, would indicate a giraffe. On the basis

of E, you think that there is a giraffe in front of you. (*) amounts to the claim: there is no *entailment* of the form *given E, it follows that there is a giraffe in front of me.*

(*) is a form of the claim that, no matter evidence we have at our disposal – no matter what is going in one's Cartesian theatre -- there is no analytic connection between that evidence and the existence of the things that we posit on the basis of that evidence. *That* claim is itself analytic. It is analytic that there is no entailment from E (your giraffe-type perception) to *there is a giraffe in the vicinity.* Statements of the form:

given such and such, it analytically follows that thus and such

and, therefore,

given such and such, it does not analytically follow that thus and such

are obviously analytic. If P is analytic, it is analytic that P is analytic. (*) is just another statement of that form; it registers the absence of an analytic connection; therefore it is itself an analytic claim. So the true modal claim conveyed by (not semantically encoded in) "Hesperus might not have been Phosphorous" turns out to be an analytic, not an a posteriori (empirical), necessity.

The statement "Hesperus is Phosphorous" certainly *conveys* some non-analytic proposition. What is that proposition? For the reasons just given, that claim cannot be: *Hesperus is Phosphorous.* For *that* claim is trivial or absurd. It must be a claim about our epistemic situation, a claim similar to (*). It is clear what, at least in its rough outlines, that claim must be:

(**) Let $H_1 \ldots H_n$ be those perceptions of yours that relate, in a certain way, to the presence of a unique celestial body that disappears after all others from the morning sky. Let $P_1 \ldots P_n$ be those perceptions of yours that relate, in a comparable way, to the presence of a unique celestial body to appear before all others in the evening sky. There is some X such that X caused both $H_1 \ldots H_n$ and $P_1 \ldots P_n$.

So the true, non-trivial claim communicated by "Hesperus is Phosphorous' turns out to be an existence claim. At the same time, the proposition *literally meant* by "Hesperus is Phosphorous" is a bare, analytic identity.

So everywhere we turn, we discover either necessary analytic truth or contingent a posteriori truth. "Hesperus is Phosphorous" turns out *not* to be anything other than trivial (analytically true) or absurd (analytically false); what *does* turn out to be non-analytic and true is some existence claim. But that existence claim is not necessary; there are -- *vide* (*) – worlds where it is false.

The above cited passages from Kripke have induced some to hold that, at the level of literal meaning, some kind of descriptivism is true of "Hesperus" and "Phosphorous". But in light of Kripke's arguments, it isn't very promising to try to re-descriptivize those names. Given Kripke's arguments. it does seem pretty hard to believe that, at the level of semantics, they are anything other than mere labels. But, for reasons earlier, our *grasp* of literal meaning is descriptive. The intuitions that neo-descriptivists want to accommodate *are* accommodated by distinguishing literal meaning from grasped meaning.

The concepts of recognition and re-identification

Whenever you learn what a term refers to, it is either through an ostensive or a descriptive definition. You are told: "*that* is Smith" or "there is some x such that [INSERT UNIQUELY INDIVIDUATING PREDICATE] and "Smith" names x." Descriptive definitions fix the referent of expressions by means existence claims. We've seen how this accounts for the difference in cognitive value between "…Cicero…" and "…Tully…"

Ostensive definitions operate through sense-perception. According to the story I tell, the information encoded in any sense-perception is existential. So whenever you see, or otherwise sense-perceive, an object, you are being given an existence-claim. A consequence is that ostensive definition collapses into descriptive definition.

There is an obvious, apparent problem with our analysis.

Suppose I met Sam twenty years ago. On your analysis, the content of my initial sense-perception of him was existential information. But I have absolutely not the slightest recollection of that information; I may have no memory of the circumstances under which I first met him. So that existential information has completely dropped out. It seems, then, that existential information is needed to initiate a concept of a thing, but not to sustain it.

This point ramifies, and its ramifications devastate your analysis. I may be introduced to the name "Tully" through some existence claim. But I may well have no recollection what that existence claim was. The same is true (mutatis mutandis) of "Cicero". Since I needn't have any recollection of the existence claims through which I was introduced to those terms, it is possible to use your strategy to explain why "…Cicero…" may differ in cognitive value from "…Tully…"

The entire foundation of your analysis is that one always grasps objects, and therewith literal meaning, through existence claims of some kind. "Hesperus is lovely" and "Phosphorus is lovely" communicate different existence claims. That is why they differ in cognitive value, even though they have the same literal meaning.

Suppose one day Pierre looks at London – let p be that perception -- and he thinks to himself: that is lovely. Some other day, he looks at London -- let p be that perception -- and he thinks to himself: that is hideous. Both that's refer to London. So, it would seem, Pierre believes both London is lovely and London is hideous. Nonetheless, Pierre is not guilty of incoherence. Your explanation is this:*

"The content of p is an existence claim (not something of the form alpha has phi) The content of p is some other existence claim (not something of the form alpha has phi). And the one existence claim does not entail the negation of the other."*

So the essence of your semantics and also of your epistemology is that information about the external world, and thus about the referents of our words, is given to us through existential information, through existence claims.

One can forget any such existential information and still have a concept of the external object in question. So those existence claims are not

essential to our concepts of external objects. They may be needed to start up those concepts. But then they can drop away, like scaffolding after a building has been completed.

The objector is right on one thing. The "start-up" existential information certainly does drop out. I may not remember, even sub-consciously, the circumstances under which I met Sam. But this absolutely does *not* mean that, at any point, I have some kind of "pure" conception of the person, i.e. one that is not paved with existential information. Indeed, if such "pure" conceptions were possible, it is hard to see how it could be possible for Cicero/Tully (Hesperus/Phosphorous) cases to arise. But, I fully grant, the objector's point definitely requires that we supplement our analysis.

Suppose my first encounter with Sam is through existential information like: *there is guy wearing a green shirt and a bowler hat, whom I am now (on occasion t) meeting and…*and, consequently, that such information is, at least initially, implicated in my interpretetings of "Sam".

I meet Sam again, four days later. And I do so with knowledge: I *recognize* him. What is happening here? My recognizing him obviously doesn't consist in my becoming aware of some bare identity; it doesn't consist in its being the case, that for some object O, I suddenly become aware that *O is identical with O.* No doubt we do recognize the truth of such bare identities. But that is not what *recognition* consists in. Recognition is not so informationally impoverished.

The story I would tell is this. I realize that

(E) there was some chap whom I met a few days earlier, on occasion t, who had such and such features (e.g. he was wearing a bowler hat)…

And I also realize that

(E) right now, on occasion t*, there is some chap, who has thus and such properties (e.g. he is wearing a baseball cap)…and I also realize that there is some one individual who uniquely satisfies both of these existential propositions.*

We see that recognizing a thing consists in seeing that some one thing uniquely satisfies each of *two* existence claims. This means that E -- the "start-up" existential information -- can be dropped. My *recognizing* Sam involves my seeing that some one entity satisfies both E and E*. Having made this recognition, I am at liberty to drop E. My awareness of Sam will be sustained, for the time being, by my knowledge of E*. And, in its turn, E* may be dropped and replaced by a third existence claim.

So what sustains a conception of an entity is knowledge of a series of interlocking existence claims. Any given one of these claims may be dropped, and probably will be. But at any juncture in the career of one's having a conception of a thing, that conception involves knowledge of *an* existence claim. And this point, as we've seen, enables us to account for the cognitive difference between "Hesperus is lovely" and "Phosphorous is lovely", or between "Cicero was wise" and "Tully was wise".

Let us deal with another objection:

You have ex nihilo produced this strange theory of conception as being sustained by knowledge of interlocking existence claims. And you claim that recognition consists in seeing that some one object satisfies two existence claims. But all of this seems very ad hoc – something you've confabulated to get your bankrupt theory off the hook.

I think that this comment has force only to the extent that we've been jaded by wrong theories of sense-perception and, more generally, of objectual awareness.

Suppose you are seeing some person named Bob. When you see Bob, embedded in your sense-perception is existential information. You cannot get rid of that existential information without eliminating the perception. This suggests that that existential information is constitutive of the content of the perception. Granted, Bob himself is (part of) the representational content of that perception. But he is given to you via this existential information.

So when you see Bob on some occasion t, the content of that perception is given by some bit of existential information like: *there is some*

chap whom I'm now seeing, wearing a bowler hat…

You see Bob again, on occasion t*. Again, the content of your perception is given by some existence claim: *there is some chap I'm seeing, wearing a baseball cap…*

On both occasions, the content of your perception is existential.

Suppose that, on t*, you *recognize* the person you are seeing is the person you saw on t. The content of your t-perception is given by existential information, and so is the content of your t*-perception. So if, on t*, you *recognize* that you are seeing the same person you saw on t, that must consist in your realizing that the same entity satisfies both existence-claims are satisfied – that there is some one entity who satisfies both. Once a few platitudes are granted about sense-perception – that sense-perceptions have existential content -- our theory of conception and recognition proves to be far from *ad hoc.*

I should now deal with what is probably the most important objection of all to our analysis: the objection that is epistemological underpinnings are bankrupt.

You say that

(14) "Hesperus is lovely"

and

(20) "Phosphorous is lovely"

are synonymous but differ in cognitive value. That is defensible, probably even true. But your explanation of why they differ in cognitive value is dubious at best. It seems to involve a false epistemology: one that makes our awareness of things less direct than it is.

Let us review your analysis. You say that one makes one's way from a token of "Hesperus" to O by way of some existence claim – one that corresponds to the peculiarities of the circumstances under which one encounters that expression. You make a corresponding claim about "Phosphorous". Starting from these assumptions, you arrive at the conclusion

that what (14) communicates to one is some existence claim, and that what (20) to one is (or may be) some other existence claim. In this way, you explain the cognitive difference between those two sentences, while conceding their synonymy.

But this story is inconsistent with the fact that we are aware of objects. When you first see Venus (Hesperus/Phosphorous), you are told "That is Hesperus". Under those circumstances, your learning the meaning of "Hesperus" involves your seeing (or otherwise perceiving) Venus. So Venus itself is an object of your visual perception. And you learn that that entity, the one we just mentioned, is the semantic content of "Hesperus". In the story you told, your learning what "Phosphorous" means involves your having a perception of that entity. You see Phosphorous: that very entity – Venus/Hesperus/Phosphorous -- is an object of your awareness. And when you are told "that is Phosphorous", you are learning that that very entity -- Venus/Hesperus/Phosphorous -- is called "Phosphorous". There are no intermediaries; there is no veil of perception. A fortiori no existential propositions serve as intermediaries: you are, in both cases, aware of Venus/Hesperus/Phosphorous. In the one case, you learn that that entity is called "Hesperus"; in the other you learn that that entity is called "Phosphorous". And that entity is safely ensconced in your awareness: you don't need to make your way to it via any existential, or other, propositions. So, it seems to me, that the epistemological conceit underlying your analysis is quite false. None of these mediating existential propositions exist; our cognitive access to objects, and therewith to semantic content, is much more direct than your Berkeleyan epistemology suggests.

If the story told by the objector were true, it is hard to see how "Hesperus is lovely" and "Phosphorous is lovely" could possibly differ in cognitive value. For, if the objector is right, our awareness of objects is so immediate and unremitting, and our ability to access semantic content so uninhibited and unconditional, that one could scarcely have any idea what "Hesperus" and "Phosphorous" meant without *ipso facto* knowing them to be co-referential.

Independently of that, the objector's point is based on a spurious epistemology. Obviously we can be aware of objects - of dogs, planets, and so forth. Barry Stroud[128] made the profound point that perception is

necessarily "predicational". You don't *just* see Fido; you see Fido *as* having these or those properties. In seeing an object, you necessarily "predicate" certain properties of it. I submit that once this is granted, the way is cleared for the analysis given above. Permit me to defend and delineate this view.

We see objects: that much is given. You see Fido: that is a given. But you don't *just* see Fido. You see a creature that occupies a certain part of space, and time; that has floppy ears and a wet nose; that appears to be under-nourished or over-nourished; that is chasing a certain other object. When you see Fido, what is given to you is not *just* Fido. What is given to you is the presence of *an* entity occupying a certain location, moving about certain ways, chasing a certain other entity…What is given to you is existential information: *there is an entity occupying a certain location, moving about in certain ways, chasing a certain other entity…*

A perception that was *just* a perception of Fido would be a surd. Whenever Fido is given to you, he is embedded, so to speak, in existential information like that just described. That existential content is obviously not incidental to your perceptual awareness of Fido: there is no way to get rid of it without getting rid of the perception of Fido. Given this, it is not entirely easy to see how that existential content could fail to *constitute* the content of your perception of Fido. And if, under the circumstances described, you were told, while having these perceptions, "*that* is Fido" or (more pedantically) "our label for that thing is 'Fido'", then that content would figure into your subsequent cognitive reactions to tokens of "Fido", presumably in the manner described above. Thus, far from being absurd, the epistemological conceit underlying my analysis seems *de rigueur*. If we deny it, we come perilously close to saying – in fact, we *do* say – that one can *just* see Fido, without seeing *a creature, in a certain place, with such and such other properties…* But one cannot see an object – an object cannot be perceptually given to one – without one's being given all of this existential content.

Of course, this thesis about perception is not an innocuous one. Some will perhaps deny it.[129] But, as epistemological theses go, it is a pretty hearty one. And once it is granted, not much else is needed to generate the analysis given above concerning the cognitive differences between "Hesperus is lovely" and "Phosphorous is lovely".

Chapter 21 Uniquely individuating descriptions and the problem of incomplete knowledge

We have yet to address the most important objection to what we've said. I've said that cognitive access to an object O is mediated through knowledge of existential-descriptive knowledge that *applies* to O. I grant that the sentence "…Gödel…" is *de re* about Gödel – that the semantic contribution of "Gödel" to a sentence containing that term is *Gödel himself*, not some concept or sense that applies to Gödel, and not some quantifier-description that turns the sentence in question into an existence-claim that is *satisfied* by Gödel. So up to this point, what I say is on the right side of Kripkean semantics.

But I *also* say that one cannot grasp Gödel himself *except* through existential-descriptive information. Of course, to grasp the semantics of "…Gödel…" one must grasp each of the constituents of the corresponding proposition; so one must grasp Gödel. If, as I maintain, one cannot grasp Gödel except through some existential-descriptive claim, then what "…Gödel…" communicates is something along the lines of: *someone x uniquely had phi and…x…*

Here is the problem. Kripke (1972) famously argued (or *seemed* to argue: see below) that people don't typically know "uniquely individuating descriptions" of the objects to which they refer and about which they think. Practically everyone has referred to Socrates and had thoughts about him. But if asked to produce a description that uniquely applied to Socrates, nine out of ten people couldn't do so. They would be able to produce *some* description, but not that one applied to Socrates to the exclusion of everyone else, and perhaps not even one that applied to him at all. They might say "he was the author of the Republic" or "he was a famous philosopher of antiquity".

This certainly *seems* to show that knowledge of a description that singles out x is *not* necessary for thinking about x. Smith may be able to think about Socrates and refer to him; but the only *description* Smith can associate with Socrates is some faulty or overly inclusive one ("author of the Republic", "philosopher of yesteryear").

This consideration induced Kripke to look for an alternative. One is able to refer to Socrates, and to cognitively access the man, *not* by virtue of knowing some individuating description, but by virtue of having a certain *causal* connection to the man.[130]

The main idea is that a *causal*, rather than a descriptive, connection to x is what underwrites the ability to refer to, and think about, x.

This idea seems to have a prototype in the fact that to *sense-perceive* an object x is not (merely) to have a mental image that "fits" x, but is to have a mental image that has the right *causal* connection to x.

The truth is that nothing Kripke says requires us to jettison the idea that cognitive access and reference are mediated through uniquely individuating descriptions. In fact, his so-called "causal" theory of reference *requires* that people know of such descriptions.[131]

What is the theory of reference that Kripke outlines?[132] Let's say I don't know of any uniquely individuating description of Socrates. Still, I can refer to and think about the man. How? Somebody encounters Socrates for the first time. (Presumably this is when Socrates is born, but that doesn't matter in this context.) That person dubs his new acquaintance "Socrates".[133] Socrates goes about, introduces himself. Some people have the good fortune to *personally* see Socrates and be told that he is called "Socrates".

Those people obviously can refer to Socrates. And they can *transmit* that ability to others. They can say: *I knew this fellow named "Socrates". He was the most obstreperous person I ever met.*…Having heard (or even just over-heard) this, the auditor of that statement can *himself* refer to, and think, about Socrates. This kind of transmission can go on ad infinitum. 2,500 years after Socrates' death, we can refer to him, thanks to a long chain of such transmissions.

Now let us re-tell this same story. But this time, let us supplement it with the epistemological points we've made.

Let us start with the people who personally *saw* (or sense-perceived) Socrates, and were told: *that* guy is named "Socrates". Notice that *those* people associate Socrates with some uniquely individuating description (not the same description in each case). Each of *those* people has a uniquely individuating description of Socrates (possibly a different one in each case). Each one of these privileged few is told something like: "Socrates" denotes *the unique person **over there**, next to the hearth, talking about the One, wearing a purple toga.*…Some of these people might actually meet Socrates. In that case, once again, one has a uniquely individuating description. Each of these people learns the semantics of "Socrates" through a description (reference-fixing, not meaning giving) like: "Socrates" denotes *the person who **I** shook hands with at time t, place p, who irritated **me** by contesting my definition of piety…*

Suppose that Larry is one of the people who were given an *ostensive*

definition of "Socrates". So Larry *saw* Socrates, possibly met him, and was told: ***that*** guy, over there…is called "Socrates".

Larry's uniquely individuating description almost certainly has an *indexical* component (indexicals are put in boldface). Larry is told that "Socrates" refers to ***that*** guy. Larry *sees* (or otherwise sense-perceives) Socrates: So Socrates is given to Larry through information like (I am speaking from Larry's perspective): *the guy who **I**'m talking to right **now*** or *the guy ten feet from **me**...*

But that doesn't in any way mean that Larry doesn't have a uniquely individuating description of Socrates. He *does* have such a description. Many uniquely individuating descriptions contain an indexical component. (In fact, it is arguable that *all* such descriptions do, except when the description in question refers to some abstract object, like the number two.)

Let us continue the story. Larry outlives Socrates. Talking to his buddy Fred, Larry says (in Ancient Greek): *I knew this guy named "Socrates", who did such and such.…*

As a result, *Fred* has a uniquely individuating description of "Socrates", namely: *the guy to whom Larry was referring on occasion C (at place p, time t).…*

A couple of points are in order. First of all, there is no circularity here. Larry had a uniquely individuating description of Socrates. That is beyond doubt. Now Larry's description was not circular: it did not have the form: "Socrates" names the guy called "Socrates". Larry's uniquely individuating description (uniquely individuating description) was something like: "Socrates" refers to *that* guy – the one next to the hearth in the purple toga…

Now Fred's uniquely individuating description of Socrates isn't circular either. Fred's uniquely individuating description is something like: "Socrates" denotes *the guy to whom Larry was referring on such and such occasion.*…Fred's uniquely individuating description is parasitic on Larry's. But Larry's isn't circular; and neither is Fred's.

Another point should be made. Larry could introduce Socrates through a highly impoverished description: *But even so, Larry has a uniquely individuating description of Socrates.* This point is crucial and has, so far as I know, always been overlooked.

Suppose Larry says: *once I met this chap named "Socrates". He was a real jerk. I wish I'd never met him. I'm sorry I brought him up. Let's not*

talk about him anymore.

It might seem at first that, under the circumstances, that Fred doesn't have uniquely individuating description of Socrates – that he has only an overly inclusive description: *some guy who Larry met who was a jerk.*

But that is not so. Fred *does* have a uniquely individuating description : *the guy who Larry was referring to at place P, time T.*…For there was *one* guy to whom Larry was referring on that particular occasion – on the occasion when Fred first heard the name "Socrates". So the description through which the semantics of "Socrates" is given to Fred (given in the reference-fixing, not the meaning-giving, sense) *is* uniquely individuating and accurate, namely: *the guy who Larry was referring to at place p, time t.…*

Further, there is no way that the ability to refer to Socrates could be transferred from Larry to Fred (or from *any* person to another) *except* by way of a uniquely individuating description (I mean a description that *in fact* singles out the relevant object). Transference of that power, from Larry to Fred, involves, at the very least, Larry's saying something about Socrates to Fred, or in Fred's presence. And no matter how cursory that something is, Fred *ipso facto* has a uniquely individuating description of Socrates. Suppose all Fred hears is Ancient Greek (or Ancient Turkish…) translation of:

(*) "that Socrates was a real nuisance. I wish he'd just stayed in Athens."

In that case, Fred knows that "Socrates", or whatever translation thereof occurs in (*), refers to the guy to whom Larry referred on that particular occasion.[134] The other side of the coin is that if Fred has *no* uniquely individuating description of Socrates, then he *isn't* connected to Socrates in the kind of way he'd have to be to refer to him or have (de re) thoughts about him. The amount of information needed to get a uniquely individuating description of Socrates is minimal. One need only hear a passing reference to the man. (Of course, seeing Socrates will do.) For then one has the uniquely individuating description : *the guy referred to on occasion C* (or *the guy I saw next to the Oracle of Delphi, gesticulating wildly while talking about the nature of piety.…*) So if one doesn't have a uniquely individuating description of Socrates, then one doesn't even have the basis for the minimal connection needed to refer to Socrates or to think about him (in a de re way).

The main point is this. Larry transfers certain abilities to Fred: the ability to use "Socrates" to refer to Socrates, and the ability to think about Socrates in a de re way. Larry's transferring these abilities to Fred involves Fred's acquiring a uniquely individuating description of Socrates.

Of course, we can imagine Fred transmitting the powers in question to some third party – say, William. By the logic we just discussed, William's acquiring these powers involves his (William's) acquiring a uniquely individuating description of Socrates. And so on, until the chain we've described reaches you, or your grandchildren.

It must be pointed out that the only legitimate uniquely individuating description that Fred has is one like: *the guy to whom who Larry referred on such and such occasion.*…Beyond that, Fred's beliefs about Socrates may be quite spurious. Fred may believe that Socrates was a person of low morality, that Socrates wrote the *Posterior Analytics*, and so on. By the same logic, the only uniquely individuating description one of us – twenty-five hundred years later – has of Socrates may be one like: *the guy who Frank was referring to that time when.*…Beyond that, one's beliefs about Socrates may be spurious.

So one *does* have to have a uniquely individuating description of Socrates to think about, or refer to, the man. But that uniquely individuating description may be highly specific to one's own circumstances, and may thus be of no interest to historians or anyone else who was interested in Socrates' biography. Russell – *rightly*, as I think we are seeing – thought that to think about Socrates one had to have a uniquely individuating description of the man.[135] But he *wrongly* took the uniquely individuating descriptions in question to be historically pregnant ones: *the guy who invented the theory of forms and drank hemlock.* He was wrong about the nature of those uniquely individuating description, but right, apparently, that one must have *some* uniquely individuating description.[136]

Thus it is no mystery that one can think of Einstein as *the guy who invented the atom bomb*. For to be able to think about and refer to Einstein, it is necessary to have *some* uniquely individuating description of the man. But that uniquely individuating description may be of no historical moment; it may be, and probably is, something like: *the guy who so and so was talking about on such and such occasion.*…That is why one can be so wrong about what Einstein did – about whether he invented the atom bomb, for example – *and still refer to and think about Einstein.*

Notice, also, that this account involves a *causal* connection between Socrates and between people today who think about and refer to him. In other words, the account we've given *is* a causal account, even though it is also a descriptivist account. Larry *sees* Socrates: obviously there is a causal connection there. That is how Larry acquires *his* uniquely individuating description of Socrates. Fred sees or hears Larry refer to Socrates on some particular occasion: obviously there is a causal connection there. And so on, until this chain reaches you. So you *are* causally connected to Socrates: and it *is* this causal connection that enables you to refer to and think about Socrates. But this causal connection is embedded in various uniquely individuating descriptions.

Clearly not just *any* causal connection between a person and a thing enables that person to refer to that thing. The causal connection operates *by way of* some conception. I am directly causally connected to many individual atoms (which bounce off many surfaces). But given some one of those atoms – call it "Alpha" – can I refer to Alpha? Is my being hit by Alpha *enough* for me to refer to it? It seems not. I have to have some way of *singling out* Alpha in thought. That way of singling out Alpha may – indeed probably will – exploit my causal connection to Alpha. But it is not the causal connection *by itself* that enables me to think about or refer to Alpha. It is some conception which (probably) *uses* that causal connection.

These points map onto, and derive considerable support from, the case of sense-perception. To see the sun, I must indeed be causally connected to the sun. But *how* does that causal connection give me cognitive (visual) access to the sun? It cognitively connects to the sun *through the indexicals in a certain piece of existential-descriptive information.* As we've seen, the content of your visual perception is, at least, in part existential (existential-descriptive). You don't just see the sun; you see a total situation involving the sun. Your visual perception "says": *that* [indicating the sun] *is an object with such and such properties (it is luminescent…)*…Or: *over there* [indicating a certain place, above the horizon perhaps] *is a luminescent object…* The "that" and the "there" obviously involve a causal connection. And they connect you, cognitively, to the object. But they operate in the context of an informational-tableau that gives you a uniquely individuating description of the sun. The uniquely individuating description will be something like: *the unique luminescent object* ***there…***Or: *the unique luminescent celestial body* ***I*** *am* ***now*** *attending to…*

The causal connection operates *within* a uniquely individuating description: it is a *part* of a uniquely individuating description. So the causal theory of perception ("I see the sun because I am causally connected to it") and what we might call the "descriptivist" theory of perception ("I see the sun because my visual state encodes a description that *applies* to the sun") are by no means incompatible: they are both true. And the same is true of the relationship between the causal theory of reference ("I can refer to Socrates because I have a certain causal connection to the man") and the descriptivist theory of theory of reference ("I can refer to Socrates because I know of a description that singles him out"). They are not incompatible; they supplement each other; they are two parts of a single true story.

To sum up, nothing Kripke says is incompatible with *our* claim that cognitive and semantic access to an object is mediated through a uniquely individuating description. (Indeed, if I am not mistaken, Kripke's *own* account of reference *involves* the claim that some kind of uniquely individuating description of x is involved in one's referring to x. But, again, I don't want to press this controversial exegetical point.) So no doubt has been cast on our central epistemological contention: one *grasps* things through existential-descriptive information. This epistemological claim not only does *not* contradict, but actually dovetails with, the contemporary epistemological point that de re apprehensions of a thing often involve a causal connection with that thing.

One clarificatory point should be made. Suppose that the only uniquely individuating description one has of Socrates, and that Phil associates with "Socrates", is: *the guy who Bill was referring to on such and such occasion.*…This fact in no way entails that "Socrates" means: *the guy who Bill was referring to.*…That uniquely individuating description has *nothing* to do with the semantics of "Socrates". It has to do with Phil's *access* to that semantics. The semantics of "Socrates" is just Socrates. (More exactly, the semantic contribution of "Socrates" to a sentence is Socrates.) Phil *accesses* that semantics – he *accesses* Socrates – through the uniquely individuating description just mentioned. To sum up, that uniquely individuating description is not part of the *semantics* of Socrates, which is the same for everybody; that uniquely individuating description has to do only with Phil's cognitive access to that semantics, which embodies the peculiarities of Phil's relationship to Socrates.

The role of uniquely individuating descriptions in our concepts

Throughout this work, we've repeatedly seen the need to distinguish between literal meaning and cognitive significance. A token of "Socrates snored" literally means: *Socrates snored.* But the cognitive significance of such a token is always a much richer proposition.

It has long been recognized that we can *refer* to objects to which we couldn't possibly have any causal connection. Many a sentence has for its literal meaning a proposition of the form:…*O*…, where O is something to which it would not be possible, even in principle, to have any causal connection.

We've already considered one such sentence. Let "Alphie" semantically contribute that individual, whosever it turns out to be, that is born before anyone else in the 3rd millennium A.D. In that case, for some object O,

(*) "Alphie will live in a densely populated world"

encodes the proposition:

(**) *O* will live in a densely populated world.

It has been said, over and over, that we can never *grasp* (**); we cannot *grasp* the proposition semantically encoded in (*). So even though (*) has a literal meaning, we cannot grasp that literal meaning. The reason, we are told, is that to have a concept of an individual, one must be *causally connected to it.*

This point involves a radical misunderstanding of the nature of our concepts -- specifically, of the role that causal liaisons between ourselves and objects have in our concepts of those individuals.

In some cases, one thinks about an object by way of a causal connection to it. But, as we just saw, the causal connection is embedded *within* knowledge of a uniquely individuating description of that object. In *every* case where we think about a spatiotemporal object, it is *entirely* in virtue of knowing a uniquely individuating description of that object. When I think about Socrates, it is *entirely* in virtue of my knowing a uniquely individuating description of that person. The descriptions need not be one of

historical moment: it need not be one like *the man who figured as the protagonist in most of Plato's dialogues.* It might be a pedestrian, historically insignificant one like: *the guy who Fred was referring to that day when he and I were having cognac.* But our concepts of external objects *always* involve knowledge of descriptions that single them out.

Thus, there is no *relevant* difference between the way in which we think about the referent of "Socrates", on the one hand, and the way we think about the referent of "Alphie" or "Julius", on the other. In *each* case, the referent is grasped, if at all, by way of a uniquely individuating description. In each case, in so far as we are able to think about the referent at all, it is by knowing some concept that applies to that thing and that thing alone.

These epistemological points are exactly parallel to the semantic points we made earlier in connection with the causal theory of reference. Reference is *always* secured by means of a description. The description is given wide-scope; and that is why the expression being given meaning refers directly to an object, and not to a concept or function that applies uniquely to that object. The causal theory, we observed, is sheer nonsense if taken as a theory as to what *constitutes* reference. What *constitutes* reference is a semantic rule. In so far as the causal theory is not nonsense, it merely registers the fact that the descriptions by means of which we fix reference often mention causal relations. But, as we saw, this doesn't warrant a causal theory of reference: Reference-fixing descriptions may allude to a person's wardrobe: but this doesn't warrant a wardrobe-theory of reference.

The uniquely individuating descriptions that figure in our concepts may allude to the causal liaisons holding among things. But in every case, conception of an external object consists in knowledge of a uniquely individuating description, not in one's having a certain causal relation to that object. In so far as a causal connection to an object is constitutive of one's concept, it is entirely because the relevant uniquely individuating description *mentions* such a connection. What constitutes conception is one's knowledge of the uniquely individuating description; not the causal connection itself. One's uniquely individuating description of an object may mention that person's wardrobe: *the first person to wear diamond-studded shoes.* But that obviously doesn't warrant a diamond-studded-shoe theory of conception.

Recognition as knowledge of interlocking existence claims

We have said that to have a conception of x is to know some true proposition of the form…*C.*…, where C is a uniquely individuating description of x. There are, of course, cases where a causal connection to x is essentially involved; but, we have seen, the causal connection is always embedded in knowledge of a proposition of the kind just described.

But there is an obvious problem with this view: a problem that doesn't threaten the things we've said, but must still be dealt with. I myself do not know who invented the fugue. Let O be that person, whoever it is. (To simplify discussion, let us suppose that exactly one person invented the fugue.) I certainly know *some* true proposition of the form:…*C.*…, where C is a description that applies uniquely to O. I know that the following are true:

(i) Somebody x uniquely invented the fugue and he had at least some musical ability.

(ii) Somebody x uniquely invented the fugue and he lived before the 19th century.

(iii) Somebody x uniquely invented the fugue and x invented the fugue.

Obviously knowing *some* uniquely individuating description that applies to O isn't enough to know who O is. And it therefore isn't enough to know which proposition is semantically encoded in sentences like:

(iv) "O was a talented composer".

(v) "O died of scurvy."

(vi) "O incurred the wrath of his less talented contemporaries."

Remember what we said earlier about "Julius" and "Newman I". Unless one knows *which* individuals these terms refer to, one doesn't know which propositions are encoded in "Julius was tall" and "Newman 1 will be a genius"; one has a simulated, but not an actual, understanding of what is meant by such sentences. One has knowledge of a *description* of the proposition which is encoded in such a sentence, but one doesn't know which proposition is thus described. One knows that "Julius was tall" is true exactly if somebody x uniquely invented the zipper and x was tall. So it is clear that

knowing who so and so is, and thus knowing the literal meanings of terms that refer to so and so, does *not* consist (merely) in knowing some uniquely individuating description that applies to so and so.

What else is needed? The short, very approximate answer is: To identify is to *re*-identify. To know who so and so is, one must know of *two* existence-claims that are uniquely verified by so and so and, further, one must know *that* some one person uniquely verifies them. Suppose you meet Bob on Monday. Since you have just met him, there is a sense in which you don't know who he is. If somebody asks you, "do you know who this is?", you must say: "I am afraid I do not." But, of course, in meeting him, you are uploading information that enables you to think about him – you are acquiring a concept of him. As we saw, this information is existential. Suppose that, a day later, you see Bob again. This time, there is a sense in which you know who you are seeing. As we saw earlier, your *recognizing* Bob consists in your knowing *two* existence claims of which he is the sole verifier, and also in your knowing *that* he uniquely verifies those propositions.

That is the *short* answer to the question "what is it to know who somebody (or something) is?". Here is the long answer. Identification is a contextual notion; it is always *relative* to some background question, or relative to some body of knowledge, that one knows, or fails to know, who somebody is. Suppose you have been living next door to Bob for years. You know very little about Bob. He doesn't talk much. He politely waves at you when he mows his lawn. But beyond that, he is a mystery. Later you and your wife are at the beach. You see somebody in the distance. Mary, who has poor vision, asks: "do you know who that is?" You, who have better vision, say: "yes: it is Bob, our neighbor." From some viewpoint, you know who it is that you are seeing: you can identify that person.

The next day, C.I.A. agents come to your house. They brusquely ask you: "Do you have any idea who your neighbor is?" You say: "I don't understand. He's just some guy who waves at me when he mows the lawn." They say: "Your neighbor is really the ringleader of an extensive terrorist organization. The government has been looking for him for years…"

When you saw Bob at the beach, you *did* know who he was, at least from one viewpoint. But, from some other viewpoint, you have *never* known who Bob is, at least not until you were briefed by the C.I.A. agents. So you did, and you did not, know who Bob was.

What happened here? As we saw earlier, having a conception of an external world object involves two things: having knowledge of existence claims of which that object is the sole verifier; having a certain kind of knowledge *about* those existence claims – more specifically, knowing that some one entity satisfies them all.

Whenever somebody verbally, or ostensively, indicates an object, some existence claim, of which that object is the unique verifier, is implicated. Mary points out Bob to you. You see Bob: in seeing Bob, what is given to you is some existence claim like: *there is somebody x to whom I am attending, and x is standing in such and such place, and has such and such appearance…*When Mary asks you "do you know who that is?", what is really being asked is something like: *there is somebody x over there, to whom we are attending, and x has such and such features: do you know who x is?* And in being asked *do you know who x is?* you are being asked to identify a more comprehensive set of interlocking existence claims of which x is the unique verifier. Because your vision is good, you are able to do this: you are able to say something like *somebody y is uniquely our neighbor and x is identical with y.* (Of course, these are not the words you would use. But this would be the import of your statement.) In saying "that is Bob, our neighbor", you are conveying (though this isn't what your words literally mean):

(*) *somebody x uniquely lives in the house to the left of ours, mows his lawn frequently, goes by "Bob", and THAT PERSON over there is identical with x.*

The only qualification is the person who you are now seeing on the beach – the person whom you can see sharply, and whom Mary can see only blurrily – is given to you through an existence claim: *there is somebody y to whom we are now attending.…*So (*) really amounts to this:

(**)*somebody x uniquely lives on the left side of our house, mows his lawn frequently, goes by "Bob", there is somebody y (over there, building a sand-castle) to whom we are now attending, and y is identical with x.*

Your background knowledge of Bob consists in your having knowledge of a set of interlocking existence claims. So your being able to identify the man in the beach consists in your being able to generate

knowledge of another such existence claim, and in your knowing of that new existence claim that it is satisfied by the same individual who satisfies the others (i.e. in knowing some one individual uniquely satisfies that new existence claim *and* the other ones). So *relative* to that background knowledge – to your knowledge of those existence claims – you know the answer to the question "who is that?"

But relative to other bodies of background knowledge, this is not the case. You know that there were various bombings, in various cities. You know that somebody or other headed the organization responsible. So you know something like:

(^) Somebody x is responsible for such and such atrocities.

You also know that you have a harmless-seeming neighbor. So you know something like:

(^^) Somebody y lives in the house to the left, and y waves at me while he mows the lawn.

When the C.I.A. agents say to you "you have no idea who Bob is", the import of their words is: somebody x is your neighbor, and somebody y committed such and such atrocities, and you don't know that x=y. *Relative* to (^) you don't know the identify of the man who waves at you as he mows his lawn. But *relative* to some other existence claim, you *do* know the identity of the person whom you and Mary are seeing at the beach. To identify someone (or something) is to connect one existence claim to some other existence claim. Identification is a relative notion. Suppose that Bob is given to you through existence claim E1. You may be able to relate E1 to E2, i.e. you may know that someone uniquely satisfies both E1 and E2. But you may not be able to relate E1 to E3, i.e. you may not know that someone uniquely satisfies both E1 and E3.

There is a sense in which any three year old knows what water is. But there is also a sense in which any three year old (except for a few prodigies) do *not* know what water is. Suppose you point to water and ask little Timmy, "what is that?" He says: "that is water." He correctly identified the substance in the glass. So, in that sense, he knows the identity of the liquid in the glass.

(w1) There is some substance x such that Timmy drinks x, bathes in x, and so forth.

Timmy knows that w1 is true. Timmy also knows that

(w2) there is some substance y such that y is in the glass that is being indicated.

Further, Timmy knows that

> (w3) There is some substance x such that Timmy drinks x, bathes in x, and so forth; there is some substance y such that y is in the glass that is being indicated; and y=x.

So *relative* to w1, Timmy can identify the substance in the glass. But relative to other existence claims, this is not the case.

(w4) There is a substance z such that z has thus and such chemical structure.

Obviously non-prodigy Timmy doesn't know that:

> (w5) There is a substance z such that z has thus and such chemical structure; and there is some substance y such that y is in the glass that is being indicated; and y=x.

So relative to (w4), Timmy doesn't know *what is in the glass.* There

is no paradox here. Identification is relative to background knowledge. Being able to identify an object is being able to connect it to background knowledge. Timmy can connect the liquid in the glass to certain bodies of knowledge, but not to others. So, from some certain viewpoints, but not all, Timmy knows *what is in the glass.*

Being able to understand

(J) "Julius was tall"

involves knowing *who* "Julius" refers to. (Here we are using "Julius" in Evans' sense, as a descriptive name that refers to that person, whoever it might be, that invented the zipper.) Knowing who "Julius" refers to involves knowing not *one* existence claim that he satisfies: it involves knowing (at least) two such claims, and also knowing that some one person satisfies them both. Further, knowing who "Julius" refers to is contextual: relative to some existence claims, you might know it; relative to others, you might not. So the concept of knowing what a term refers to, and therewith the concept of being able to assign meaning to a term, is contextual: so, therefore, is the concept of understanding a sentence-token.

A loose end: Fodor's folly

For reasons discussed earlier, Kripke's landmark work made it *seem* plausible to suppose that, if x is to refer to y, or think about y, then x must have a certain kind of causal connection to y. This position is reinforced by Putnam's classic thought experiment.

Fodor eagerly accepts the view that, for me to have a concept of x, it is *necessary* for me to be causally connected to x. But he went much further. He said that for me to have a concept of x just *is* for me to be causally connected to x. The causal connection is *necessary and sufficient.* Fodor advocates an extreme version of the causal theory of conception (CCC)

There are two versions of (CCC): the strong and the weak.

Weak version: For me to have a concept of x, it is necessary for me to be causally connected to x.

Strong version: for me to have a concept of x just is for me to be causally connected to x.

We've seen that the weak version is false. Conception is *never* identical with, or constituted by, a causal connection between subject and object. Conception is *always* constituted entirely by knowledge of a description that singles out the object. A causal connection may be *mentioned* in that description. But it is never constitutive of the concept in question. Given that the weak version is false, it follows that the strong version is false.

We spoke earlier about the problems bedeviling the causal theory of reference. We can refer to things to which we have no causal relation. Also causal relations aren't fine-grained enough: any causal relation I have to the clay is also a causal relation to the statue. But I can refer to the clay without referring to the statue (and *vice versa*).

Also, the causal theory patently doesn't apply to cases where the expression in question has constituent-structure.

Exact analogues of these problems bedevil (CCC). Fodor has produced some extraordinary solutions to these problems. I invite the reader to consider them. Many of the problems with Fodor's solutions are extremely well-documented; there is no point in repeating them here.

But there is one problem that, curiously, has not yet been mentioned. Fodor very clearly says that, where "patently phrasal" concepts are concerned – concepts that are clearly built out of other concepts – CCC does not apply. So CCC is not meant to apply to my concept *angry, brown cow.* CCC is meant to apply to my concept *angry*, my concept *brown*, and my concept *cow*: but not to concepts constructed out of them. But this, it seems to me, disjunctivizes the concept of a concept. In some cases, a concept is a causal connection. In other cases, it is not. Fodor could respond by saying:

In all cases, a concept of something is an awareness of that thing, and is thus an ability to think about that thing. In some cases, such an awareness is constituted by a causal connection; in others it is not. CCC is a statement about what constitutes concepts. Different things can constitute concepts. Concepts they are multiply realizable – just like moral and aesthetic properties.

Superficially, this is a good response. But, on examination, it proves to reveal yet another deep arbitrariness in content-externalism. [This has to be completed]

Chapter 22 Program-causes and the Jackson-Pettit Attempt to Save Externalism

As we've seen, there are some compelling *prima facie* reasons to think that certain forms of externalism strip mental content of any causal powers. If that is the case, then we have a good reason to reject that doctrine. Jackson and Pettit have produced a brilliant argument purporting to show that, in fact, externalism *is* compatible with the presumption that mental content has causal powers. Their argument involves an original and (I believe) correct point about causation. But I believe that their *application* of this correct point to this particular problem is flawed. They have not, I will argue, given us any reason to think externalism is in fact compatible with the presumption that mental content has causal powers.

Like externalism, another widely accepted doctrine about the mental – namely, functionalism – is *prima facie* guilty of stripping the mental of any causal powers. Jackson and Pettit use the aforementioned analysis of causation as a way of saving functionalism. Their attempt to save functionalism, I will argue, is no more successful than their attempt to save externalism.

§ Let us start by reviewing the basics. Smith and Twin-Smith are atom for atom duplicates. But Smith is thinking about Mary, while twin-Smith is thinking about twin-Mary. The reason is that Smith's brain-state has one distal cause (Mary), while twin-Smith's has another (twin-Mary). So, according to the externalist, this difference in distal cause makes for a difference in content.[137]

But this difference in content seems to be inert. What does all the work is *local*: the spatially and temporally local property of Smith's brain are what cause him to pick up the phone and dial those numbers; the same is true (mutatis mutandis) of Twin-Smith. The differences in distal cause are causally irrelevant. Everything that does causal work is in the here and now, and is lodged inside Smith's and twin-Smith's respective crania. It is irrelevant that Smith's state was distally caused by Mary, as opposed to twin-Mary or hologram-Mary or Robo-Mary. Smith's state might as well have

been caused by any one of them, so far as the causal properties of that state are concerned.

But if the externalist is right, then Smith and twin-Smith have qualitatively different mental *contents*, corresponding to the differences in distal causes. Smith is having a thought about Mary, while twin-Smith is having a thought about twin-Mary. So, if externalism is right, that difference in content is inert.[138] What is *not* inert – what does all the causal work – is precisely what Smith and twin-Smith have in common; and what they have in common is neutral between Mary and twin-Mary. The fact that Smith is thinking about Mary – as opposed to twin-Mary or Robo-Mary – is without causal consequences. So far as what he is doing right now, his current condition could have been caused by any one of the three. So to the extent that mental content is to be understood in externalist terms – to the extent that Smith and twin-Smith have qualitatively different mental contents in virtue of their differing causal liaisons -- mental content doesn't do anything. Externalism thus appears to strip mental content of causal power.[139]

§ Jackson and Pettit deal with this problem in a brilliant way. We must distinguish causal *efficacy* from what they refer to as causal *relevance.*

In everyday discourse, we say things like the following.

"The liquid in the glass froze (thus expanding); and that is what caused the glass to shatter."

Well, what *really* caused the glass to shatter is something much more specific; it is the fact that, in certain very specific places, certain very specific pressures were put on the walls of the container.

Consider the statement:

"The presence of a flame under the container caused the liquid to boil."

Well, what *really* caused the container to heat up is much more specific: certain specific displacements of energy led to some other specific displacements of energy; and these displacements happened in a very specific region – a region coincident with, or vanishingly close to, the region occupied by the under-surface of the pan. Not every displacement of energy constitutive of the flame caused the increase in temperature. It wasn't really *the flame* which did the work: it was certain very specific micro-events constitutive of the flame.

One last example:

"The pianist hit a wrong note because *somebody* coughed too loud."

Well, what *really* caused the pianist to hit a wrong note is much more specific. *Smith* coughed too loud – it wasn't just *somebody.*

Causes are much more specific than our statements typically indicate; and the things we describe as "causes" – *somebody's* coughing too loud, a *flame's* being underneath the container, *the liquid's* expanding – are not really what is doing the causal work. There is always something much more specific going on: some specific molecule (or atom or quark…) bouncing off of some other molecule (or atom…). In fact, in the end, the only things that are doing causal work are specific, minute displacements of mass-energy. *The flame* is much too expansive and gross an entity to be responsible for the fact that this, that, or any other specific part of the pan heats up. What heats up place x on the pan is always some specific mass-energy displacement.

But when we say things like "*someone's* coughing caused the pianist to hit a wrong note", "the presence of the flame caused the pan to heat up", and so on, what we are saying is surely not altogether wrong. There is plainly a sense in which it is true to say that *somebody's* coughing is what caused the pianist to error, a sense in which such statements are true.

Here is how Jackson and Pettit deal with it. The presence of the flame *programs for* the heating-up of the pan. Put another way, it is *causally relevant* without being *causally efficacious* in the heating of the pan. What

does this mean? Given that there is a flame present, it necessarily follows that something which *is* causally efficacious in the heating up of the pan will be present. The flame is constituted by the specific displacements which did the causal work. And the presence of a flame in that place guarantees that *some* such displacements will be present which will do the causal work. So even though the flame *per se* is not what does the work, the presence of the flame is causally *relevant* to the heating up of the pan: the presence of the flame guarantees or "programs for" the presence of something which would do said work.

The same principle (mutatis mutandis) applies in the other cases. What causes the glass to shatter is not the fact that the liquid inside has a certain temperature or volume. It is something more specific: the impact of certain molecules on certain other molecules. (Actually it is even more specific than that: it is specific mass-energy displacements constitutive of the molecules.) But the temperature of the glass is causally *relevant.* Given that the glass has such and such temperature, it is guaranteed, or "programmed for", that there will be the kind of thing which *is* causally efficacious in the shattering of the glass.

What caused Horowitz (the pianist) to hit the wrong note was *Smith's* coughing or *Brown's* coughing: some *specific* person had to cough. The general fact that *somebody or other* in the vicinity coughed isn't what did the trick. But that general fact does guarantee or program for the existence of the sort of singular fact which *would* do the trick. Given that *somebody* or other coughed, it follows that, for some specific x, x coughed: and the fact that x coughed *is* causally efficacious as regards Horowitz's erring. So *somebody or other's coughing* is causally relevant, though not causally efficacious, to Horowitz's erring. It *programs for* the existence of something which is causally efficacious, even though it does not itself have that property.

Sometimes Jackson and Pettit describe the thing which is causally *relevant*, but not causally efficacious, as the "program cause". So the fact that *somebody* coughed is the program-cause of Horowitz's erring; the fact that the flame was underneath the pan was the program cause of the pan's heating up to such and such temperature.[140]

§ Jackson and Pettit say that, if we take into account the distinction between causal efficacy and causal relevance, we can be externalists *and* hold

onto the idea that content has an important causal role. For the reason earlier given, what is causally *efficacious* in Smith's reaching for the phone and dialing those numbers is not the fact that Smith is thinking about Mary. But Smith's thinking about Mary is *causally relevant.* His doing so "programs for" the kind of thing that will do the trick – that *will* be causally efficacious in his hitting those seven numbers. So content has causal relevance (though not causal efficacy). As far what is causally *efficacious* is concerned, it is irrelevant whether the distal cause of his current condition was Mary, twin-Mary, Robo-Mary, or a hallucination. Thus Smith's thinking about Mary – as opposed to twin-Mary or Robo-Mary – is not causally efficacious. But Smith's thinking about Mary is still *causally* relevant: his thinking about Mary *programs for* his dialing those seven numbers.

The flame *programs for* the specific micro-events E1…En which are causally efficacious in the heating of the pan, even though the flame is not *itself* thus efficacious. By distinguishing causal relevance (the property of programming for something causally efficacious) from causal efficacy, we recover our intuition that the flame has an important *causal* role in the heating of the pan. Similarly, Smith's thinking about Mary *programs for* the occurrences of events E*1…E*n, which are causally efficacious as regards his dialing those seven numbers, even though his thinking about Mary – as opposed to Twin-Mary or Robo-Mary – is not thus efficacious. Once again, by distinguishing causal relevance from causal efficacy, we recover our intuition that Smith's thinking about Mary has an important causal role in Smith's thinking and acting. In general, by making that distinction, we reconcile our intuitions about the causal importance of representational content in mental and behavioral life with the evident truth of some kind of content externalism.

§ Before we deal with this formidable argument, a background point is in order. Sometimes Jackson and Pettit analyze the notion of causal relevance in counterfactual terms: x is causally relevant to the occurrence of y if, in any counterfactual circumstance where x exists, there also exists something causally effective in bringing about y. The problem is that this is too generous. My striking the glass with the hammer is sufficient, in the counterfactual sense just defined, for the breaking of the glass. Now the same is true of my striking the glass *while* wearing a yellow hat and having green teeth and wearing a spacesuit…But surely my wearing a yellow hat and

having green teeth are not in any sense "causally relevant" to the breaking of the glass. The counterfactual analysis is too broad, too expansive.

The counterfactual analysis of causal relevance needs to be tightened. But we don't want to tighten it too much. For example, we don't want to say that the program-cause is the *smallest* thing x such that x counterfactually guarantees the presence of something causally efficacious. After all, the smallest thing that gives that guarantee is the mass-energy displacement itself. But we don't want to say that *only* that mass-energy displacement is causally relevant: surely we want to allow that the flame of which it is a part is causally relevant; we don't want to shrink causal relevance down to causal efficacy. That would eviscerate the distinction between causal efficacy and causal relevance.

The notion of causal relevance is, I believe, to be understood in terms of the notion of *redundancy* – in terms of the notion of a *back-up mechanism.* Consider the case where, as we loosely say, the presence of the *flame* caused the pan to heat up. Let x be the *specific* mass-energy displacements constitutive of the fire which, in the actual case, did the trick. The flame is causally relevant because in any case where you have the flame, but where x (for some reason) didn't do the trick, the flame guarantees that some other mass-energy displacements y *will* do the trick. So causal relevance can be understood in terms of "back-up" mechanisms. In any situation where *someone or other* in the vicinity coughed loud enough, but where Smith (the actual cougher) did not cough, that is enough to make Horowitz hit the wrong key.

Now Smith's wearing a yellow hat is *not* causally relevant to Horowitz's erring, because it doesn't constitute the right kind of back-up mechanism; it doesn't amplify the range of cases where something is causally efficacious in getting Horowitz to err. By contrast, the *whole* flame is causally relevant to the heating of the pan because, if the flame were smaller, that would eliminate some back-up mechanisms; it would shrink the range of mass-energy displacements that *would* step into the breach in situations where x (the totality of mass-energy displacements that *in fact* did the trick) isn't being causally efficacious. But Smith's wearing a yellow hat, the flame's coming from a green as opposed to a blue lighter, or its coming from the lighter of somebody wearing a plaid as opposed to a solid green shirt, doesn't provide this sort of counterfactual reinforcement. It is inert in this respect.

The same is true of the fact that Smith's brain-state has its origin in Mary, as opposed to twin-Mary, or Mary-7, or Robo-Mary, or hologram-Mary or a Cartesian demon or bent light-rays. What is causally *efficacious* is some congeries X of mass-energy displacements at various contact-points in Smith's nervous system. What is causally *relevant* are the structures that guarantee, in counter-factual situations where X didn't do the trick, that some other congeries Y would. What is *not* causally relevant is anything that doesn't provide this sort of reinforcement. Where such reinforcement ends, so does causal relevance.

Now consider the things going on outside Smith's cranium. Are any of *these* things going to step into the breach in counter-factual situations where X isn't doing the trick? Consider a counter-factual situation where X isn't doing the trick. Is the fact that Smith's brain-state originates with Mary (as opposed to twin-Mary or Mary3) going to help out? Is it going to provide the needed back-up? No. It is inert. It would provide back-up *only* to the extent that it led to the right kind of *brain-structure* or, in any case, the right kind of *local* phenomenon that would provide back-up.

§ In this context, we have to make sure that our terminology isn't doing the thinking for us. Jackson and Pettit are using the term "causally relevant" in a technical sense. But, of course, that expression has a non-technical sense; one could use it in ordinary speech, and be understood. In the non-technical sense, the Big Bang is obviously "causally relevant" to everything that is now the case, including the fact that I'm typing on a computer. In fact, give or take a few niceties relating to quantum-indeterminacy, the Big Bang was causally sufficient for that, and every other, contemporary fact. But the Big Bang is not "causally relevant" in Jackson and Pettit's technical sense.

The concept that Jackson and Pettit are labeling with the expression "causal relevance" is an important, coherent, and explanatorily fecund one[141]. And that concept deserves *some* kind of label. But it is easy to conflate the technical meaning of "causally relevant" with its original, non-technical meaning. And such a conflation could easily create the illusion that the concept of causal relevance has application where it really does not. In fact, I feel that, in their application of their insights into causation to the issue under discussion, Jackson and Pettit themselves may be guilty of this sort of conflation.

Let us use the expression "causally relevant$_{JP}$" to refer to causal relevance in the technical Jackson-Pettit sense. And let us use the expression "causally relevant" to refer to causal relevance in the non-technical, pre-Jackson-Pettit sense. The Big Bang is obviously causally relevant to the fact that I am now typing– after all, it is the ultimate *cause*: and how could it be more causally relevant than that? But the Big Bang is not causally relevant$_{JP}$ to my typing. (Having made this point, let us now go back to using the term "causal relevance" – no subscript – in the technical sense in which Jackson and Pettit use it.)

Let us consider the paradigms through which Jackson and Pettit give meaning to their technical use of the term "causal relevance". Horowitz hits the wrong note. What causes that to happen is *Smith's* coughing or *Brown's* coughing: it isn't just *someone's* coughing. (Causal efficacy resides in singular facts.) But the general fact that *someone* or other coughed is causally relevant to Horowitz's hitting the wrong note . Why? Because that general fact guarantees that there will be a singular fact of the right kind; it guarantees that, for some individual x, there will be a singular fact of the form *x coughed.* Now what is the nature of this guarantee? We are not talking about a *causal* guarantee. It is not as though *someone's* coughing at time t *causes* Smith's coughing at time t+1, which in turn causes Horowitz to hit the wrong note. No – the guarantee here is much tighter than that. It is in the nature of a logical or metaphysical guarantee. Given that *someone* coughed (in the right vicinity), it follows logically – not causally – that, for some specific x, *x coughed.* So the fact that *someone* coughed logically guarantees that the right kind of singular fact will obtain.

What causes the pan to heat up is the fact that there are certain *specific* events in an area coincident with, or vanishingly close to, the bottom of the pan. The much more encompassing fact that there is a *flame* underneath the pan isn't what does the trick. What does the trick is much more specific. But the just-mentioned, more encompassing fact is causally relevant to the heating of the pan. Why? Because it *guarantees* that the right kind of specific events will occur. What is the nature of that guarantee? It isn't a *causal* guarantee. The specific mass-energy displacements which do the trick are *constitutive* of the flame. The flame comprises innumerable mass-energy displacements; and the ones which are causally efficacious in the heating of the pan are among them. So the aforementioned guarantee is not *causal.*

Here is the wrong story:

The flame exists at t, and that causes there to be certain other events E1…En at time t+1; and, in their turn, E1…En cause the pan to heat.

That isn't how it works. Here is the right story. The flame is simultaneous with E1…En. (This is to be expected, since E1…En are constituents of the flame.) The flame guarantees the existence of the right kind of specific pan-heating events in the sense that the flame is *constituted* by all manner of events which will do the trick. If the flame were present but, for some reason, E1…En didn't occur, other events E*1…E*n *would* occur such that E*1…E*n would do the trick. It isn't that the flame would *cause* E*1…E*n to happen; it is that the flame would *comprise* those events. The flame is *constituted* by the kinds of things which will be causally efficacious in the heating of the pan. And that is the sense in which the presence of the flame underneath the pan "programs for" the heating of the pan; that is the nature of the guarantee previously mentioned. The presence of the right kind of events is guaranteed not in a causal, but a metaphysical (for lack of a better word) sense: the flame doesn't *cause* there to be the right kind of events; the flame is *constituted by* the right kind of events.

Exactly similar remarks apply to the case of the liquids shattering the walls of the vessel. Here is the wrong picture:

The liquid expands at t. That expansion *causes* certain other events E1…En. E1…En, in their turn, cause the walls to crack.

Here is the right picture. The liquid expands at t. That expansion *involves* the occurrence of certain events E1…En. And E1…En (the specific mass-energy displacements at, or vanishingly close to, the interior surface of the vessel) cause the vessel to crack. If the liquid-expands, among the events *involved* in that expansion – among the events constitutive of it – are such as will be causally efficacious in the cracking of the glass. So it is not in a *causal* sense that the expanding of the liquid guarantees the existence of causally efficacious events. The guarantee is metaphysical (for lack of a better word): the expansion is *constituted* by the right kind of events.

Let us bring these points to bear on the Mary/twin-Mary case. Smith decides to hit seven numbers on his phone (so as to call Mary). Mary -- or,

better, some state of affairs involving Mary -- is a (contributory) distal cause of this. Let S1 be that state of affairs, that distal (contributory) cause. And let S2 be the event of Smith's hitting those seven numbers. It may be that the occurrence of S1 in some sense "guarantees" the occurrences of S2. But the guarantee here is causal. S1 is buried in the remote recesses of space-time. (To fix our ideas, suppose that Mary and Smith haven't seen each other in sixty years.) There is no sense in which S2 is constitutive of S1 (or vice versa). And, of course, there is no purely logical relationship between them. So the sense in which S1 guarantees S2 isn't remotely comparable to the sense in which *someone's* coughing guarantees that, for some specific x, *x coughs*; and it isn't remotely comparable to the sense in which the presence of the flame guarantees the existence of the right kind of specific mass-energy displacements at, or vanishingly near, the under-surface of the pan.

Now *there are* events which, while not themselves being causally efficacious in making Smith dial those seven numbers, do yet *guarantee* that he will do so, where "guarantee" is being used to denote the right logical or metaphysical relationship – the kind that holds between someone's coughing and Smith's coughing, between the flame's being underneath the pan and the pan's being heated. What is it that is causally efficacious in making Smith dial (or at least choose to dial) those numbers? Certain specific mass-energy displacements at, or vanishingly near, certain neural contact points. So what is causally efficacious isn't exactly some chemical (or electrical) storm in Smith's brain – some flurry of z-fibers or whatnot. The *storm* is much too encompassing an entity to be causally efficacious (just as the *flame* is too encompassing an entity to be causally efficacious in heating the pan): for causal efficacy resides at (or vanishingly near) contact points; the storm *includes* the things that happen at those contact points, but it also includes things that are (relatively speaking) far away from them. Nonetheless, that chemical storm – the thing that neurologists speak of as "the cause" of Smith's number-pushing muscle-spasms – in some sense *guarantees* that the right kind of events will happen at those contact points. What is the nature of that guarantee? The chemical storm *consists* of those events (and others). The occurrence of the storm guarantees the existence of the right events at the relevant contact-points because the storm *consists* of events that will do the trick. In the actual world, the storm comprises E1…En – these being the relevant events at the contact-points. But if the storm hadn't comprised specifically E1…En, it *would* have comprised E*1…E*n, where E*1…E*n

would have done the trick. It *would* have done this because the chemical storm is *composed of* the right kind of events – not because it would have *caused* the right kind of events. So what is causally relevant to, but not causally efficacious in, Smith's dialing those numbers is *not* some long gone state of affairs involving Mary: it is some contemporaneous state of affairs which *comprises* the causally efficacious events. Events buried in the past – whether they involve Mary, twin-Mary, Robo-Mary, or that rascal Oscar who likes to impersonate Mary – are not, in Jackson and Pettit's technical sense of the phrase, "causally relevant" at all. They are *causes*; but they are not causally efficacious – for causal efficacy resides in what is doing the work right now – and they are not "causally relevant" (in the technical sense) -- for what is causally relevant *comprises* what is doing the work right now.

Notice that, in the paradigms through which Jackson and Pettit define their technical term "causal relevance", the state of affairs that is causally relevant is always contemporaneous with the state of affairs that is causally efficacious. *Someone's* coughing must happen at the same time as Smith's coughing. The flame's being underneath the pan happens at the same time as E1…En occurring at, or vanishingly near, the under-surface of the pan. The liquid's expanding happens at the same time as E*1…E*n happening at, or vanishingly near, the interior surface of the beaker. But with regard to the state of affairs involving Mary – or, better, involving long dead star Beta – in virtue of which I am now thinking about Mary as opposed to twin-Mary, about Beta as opposed to twin-Beta – that state of affairs is quite definitely *not* contemporaneous with the events which are causally efficacious in my dialing those numbers, not even after we take into account the adjustments to our concept of simultaneity demanded by relativity-theory and the like. And this should tell us something as to whether the Mary/twin-Mary case is on all fours with the Smith-someone/flame-E1…En case.

§ *Functionalism and Mental Inertness*

Here we might as well take an opportunity to discuss a problem that does not lie squarely on our path. We have a justification for this. The digression concerns an important matter – the viability (or rather lack thereof) of an almost universally accepted conception of the mental. And given the points we've just made, we are in a position to pounce on this issue.

It seems reasonable to suppose that, in some cases, brain-states

realize mental contents. (I wish the word "state" to be taken as broadly as possible – to apply to both enduring brain-structures and occurrent phenomena.) My belief that 1+1=2 is realized by certain brain-states. My belief that water is wet is realized by a different set of brainstates.

Let x be an arbitrary brain-state of mine; and suppose that x is identical with, or realizes, my belief that bread nourishes. According to functionalism, for x to realize that belief is for x to have a certain causal role. For x to have a certain causal role is for it to be caused by certain things (e.g. visual perceptions of people eating bread and subsequently appearing sated) and for x, in its turn, to cause certain other things (e.g. my choosing to eat bread when hungry). So, for functionalism, x is a belief that bread nourishes in virtue of (inter alia) the fact that x leads to my deciding to drive to the store to buy bread, to my eating said bread, to my teaching loved ones how to bake bread, and so forth.

There is an obvious problem for functionalism. If x's realizing a belief that bread nourishes *consists* in (inter alia) its leading to certain states of affairs – e.g. my choosing to eat bread, my deciding to drive to the bread-store when hungry, and so forth – then x's having that content cannot be what *causes* those states of affairs to come about. If x's having that content *is* its having a certain causal role, then x's having that content cannot *affect* or *govern* its causal role; it cannot be what brings about my choosing to eat bread or my teaching my loved ones how to bake it. In a word, if content *is* causal role, then content cannot *affect* causal role: content ends up *not* having a causal hand in contexts where, quite plainly, it *does* have such a hand.[142]

§ The nature of this problem for functionalism becomes clear if we are careful to take into account a distinction that we discussed earlier. Strictly speaking, it is not *things* that have causal powers. It is not *the rock* which breaks the window, but the rock's moving with a certain velocity, and having a certain mass, and so on. *The rock* doesn't do anything. What does the work is really the rock's having certain properties – a certain mass, solidity, velocity, and so on – along with various other things' having various properties (e.g. the window's having a certain micro-structure).

Not every one of the rock's properties is causally effective in *every* causal transaction involving the rock. Suppose the rock has a certain dimple on one of its sides. In some contexts, the rock's having that property will be causally effective. (For example, if you pour water down that side of the

rock, the presence of the dimple will affect the course of the water.) But in other contexts, the presence of that dimple will *not* be causally effective. (For example, if you drop the rock in a vacuum, the presence of the dimple won't have any effect on the trajectory of the rock.)

Suppose, once again, that x is a brain-state realizing my belief that bread nourishes. Strictly speaking, x doesn't do anything: what does the work is x's having certain properties – e.g. x's having certain electrical, chemical, and cellular properties. Not every causal transaction involving x is one where x's *electrical* properties are causally effective. Perhaps in certain contexts it is x's mass, not its electrical properties, that are causally effective. Similarly, not every causal transaction involving x is one where x's *mental* properties are causally effective. In some contexts, x's mental properties might be causally irrelevant: what is doing the work in those contexts might be x's having a certain electrical charge or its having a certain specific shape. So there is a difference between *x's* being involved in some causal transaction, on the one hand, and *x's realizing a belief about bread* being causally effective in that context. The latter might or might not be causally effective on that occasion.

But, of course, we want to believe that, in at least some cases, my believing that bread nourishes is causally effective. So we want to believe that, in certain contexts, x's *having a certain mental content* is causally effective. What leads me to eat the bread in front of me is my believing that bread nourishes. Here what is effective is not just x, but is x's having the right mental properties.

Here is where the problems for functionalism begin. The functionalist obviously allows that *x* has causal powers (it isn't in the same category as a number or a disembodied spirit). If x's being a belief that bread nourishes *consists in* its having such and such effects – e.g. its leading me to choose to eat bread, to decide to teach my children how to bake bread, and so on – then x's having that content cannot be *causative* of my choosing to eat bread, bake it, teach my children how to bake it, and so on.

Surely x's being a belief that bread nourishes is what *causes* me to choose to eat bread (when hungry), to buy bread, and to teach loved ones how to bake bread. But if x's being such a belief is *identical* with, or *consist in*, its having those effects – if it consists in its bringing about those states of affairs – then x's being such a belief cannot be a *cause* of those states of affairs. If x's being a belief that bread nourishes *consists in* x's leading me to

choose to buy bread, eat said bread, and so forth, then x's being such a belief cannot *cause* me to do those things. After all, nothing is a cause of itself.

So by *identifying* content with causal role, functionalism deprives content of the power to have any *affect* on causal role. By *identifying* content with causal role, functionalism strips the mental of ability to cause the things which we know it to cause.

Let us sum up. If content is *identified* with causal role, then content cannot be what governs causal role. But in that case, content doesn't *do* anything; it becomes inert, a free-wheel or epiphenomenon. But plainly content is *not* inert: content does a lot. So functionalism must be wrong.

§ Jackson and Pettit say that, despite what we just said, functionalism *is* consistent with the presumption that the mental has causal powers. Here is what they say.

For the reasons just given, Smith's wanting water cannot be *causally efficacious* in Smith's reaching for the glass: it cannot, strictly speaking, be what causes Smith to do that. But it can be causally *relevant.* Smith's being thirsty "programs for" his doing that sort of thing; the presence of thirst guarantees the presence of the kinds of things which would do the causal work in Smith's reaching for the glass. So when we take into account the distinction between causal efficacy and causal relevance – between *bona fide* causes and program-causes – we reconcile functionalism with the presumption that the mental is causally potent.

§ Once again, I believe that Jackson and Pettit have had a profound insight into causality and causal explanation. But I don't think that insight helps out functionalism in the way that they think.

Let us start with some platitudes. Nothing can be "causally relevant" to itself. A state of affairs cannot be "causally relevant" to its own existence. Obviously things are relevant to themselves. But the relevance isn't *causal.*

Of course, the term "causal relevance" (as Jackson and Pettit are using it) is an invented one; so we can give it any meaning we wish. But surely it would be deeply misleading to describe the kind of relevance that event E has to itself as *causal.* The relation is better described as logical, or perhaps metaphysical. E is not *causally* related to itself; and it doesn't matter whether we are taking about causal efficacy or causal relevance. Obviously

the kind of relation Jackson and Pettit have in mind, when they talk about "causal relevance" is *some* kind of causal relation; it is not a relation which something can have to itself; the word "causal" is there for a reason. So nothing can be "causally relevant" to itself.

The functionalist says that x's causing E1…En is what x's having C *consists in.* There is an identity here, not a causal relation. At the same time, it is clear, from a common-sense point of view, that x's having content C – e.g. x's being a desire for water – has some kind of causal hand in bringing about E1…En (i.e. that it has a causal hand in bringing it about that Smith reaches for the glass). If x's having C is *identical with* x's bringing about E1…En, x's having C cannot be *causally effective* in bringing about E1…En: after all, x's bringing about E1…En cannot cause x's bringing about E1…En. By the same token, x's having C cannot be *causally relevant* to the bringing about of E1…En. Nothing can be *causally relevant* to itself. If the functionalist is right, then x's bringing about E1…En is *identical* with x's having C. So if the functionalist is right, then x's having C – i.e. its bringing about E1…En – cannot be causally *relevant* to its bringing about E1…En. It cannot be causally relevant, any more than it can be causally effective. So the Jackson and Pettit maneuver fails.

Chapter 23 Externalism and the Veil of Perception[143]

The points we've been discussing throughout this book allow us to weigh in on a perennial issue in philosophy: the question whether – and, if so, how – we are aware of external objects? I would like to end this section of the book by making a few remarks on this topic.

There are two opposed conceptions about our epistemic relation to external reality. Some (e.g. Hume, Ayer) say that we perceive *representations* of external objects, and not the objects themselves: there is thus a "veil of perception" lying between ourselves and the outer world. Others (e.g. Reid) say that we perceive the objects themselves.

Sometimes the former position is known as "indirect realism": there really are external objects, but we are only indirectly aware of them – we are aware of representations of them. Predictably, the latter position is sometimes referred to as "direct realism": there really are external objects, and we are directly aware of *them*, not of representational intermediaries. (In this context,

I am taking for granted the existence of the external world; so it isn't worth discussing the relations of internalism or externalism to *anti*-realism.)

On the face of it, it might seem as though *externalism* is associated with direct realism, while internalism is associated with indirect realism. According to the externalism, Alpha Centauri itself is a veritable constituent of my thoughts. No Alpha Centauri, no thought. Thus, given that the thought exists, there is no need to *connect* the thought to external reality: if it exists at all, the thought exists *connected* to the external world.

By the same token, indirect realism seems to be associated with *internalism.* The internalist says that the content of our mental states is confined to what is internal to the "Cartesian theatre". So it would seem that, for the internalist, the content of our states does not, at least not directly, have anything to do with anything that lies outside the Cartesian theatre: Alpha Centauri is not *itself* a constituent of our thoughts. What is a constituent of our thoughts is, at most, something that *represents* Alpha Centauri. So internalism appears to be committed to the view that perceptual awareness is mediated by a veil of representations. (Let us refer to this view as VOP for "veil of perceptions.)

There is some historical tendency for internalists to accept VOP – for example, Descartes accepted VOP and was an internalist, so far as he can be classified either as an internalist or as an externalist. This is not universally true. John Searle is a direct realist who is also an internalist.

I think that this debate needs to be re-examined in light of some of the points we've discussed. Thus re-examined, it turns out that, ironically, externalism is committed to a rather extreme form of VOP. A good point of entry into this debate is "Brentano's thesis":

(BT) All awareness is awareness of something.

(BT) seems reasonable enough. It is hard to imagine a "pure" awareness – an awareness that wasn't an awareness *of* this or that.

But there is a famous problem with (BT). Suppose you hallucinate a cat. What is the object of your awareness? A cat? There is no cat – at least, no *real* cat. Yet in having the hallucination, you are obviously aware. So here, it will be said, we have an awareness that is not an awareness of anything. Thus, it seems, we have a counter-example to BT.

Brentano, and others, have dealt with this by saying that, when you

hallucinate a cat, you *are* aware of a cat: a non-existent (or merely "subsistent") cat. But this is obviously not good reasoning. Either there is a cat or there isn't. If there isn't, then there isn't any cat to be aware of. Either we are simply playing with labels or we are ontologizing to cover up deficits in our reasoning.

But hallucinations are not counter-examples to BT. For reasons earlier discussed, the content of your perception is existential. When you 'see Fido', as we put it, what your perception tells you is existential: *there is a creature with such and such properties.* The perception is *made true* by some fact about Fido. But Fido is not himself a constituent of the content of that perception. To think otherwise, we saw, is to confuse perceptual content with *meta*-perceptual content. Given *only* the information encoded in your perception, you have no way (no matter how intelligent you are) of knowing whether you are seeing Fido or twin-Fido or Android-Fido (or nothing at all). The only information that could help you decide that is meta-perceptual – information obtained through some source other than that very perception. What is given to you in that perception is neutral between Fido's being in front of you and twin-Fido's being in front of you and android-Fido's being in front of you. There is nothing *in that perception* itself that could possibly warrant a decision as to which, if any, of these three things (Fido, twin-Fido, Android-Fido) is the one in front of you. Thus the perception cannot deliver a message of the form…*Fido.*…(as in *Fido is looking ill* or *Fido is panting heavily*). For if the perception did deliver such a message, it would *not* leave it open, even epistemically, whether it was Fido or twin-Fido (or neither) that was in front of you. So Fido cannot be a constituent of the information conveyed to in you that perception.

What is given to you in that perception is Fido-neutral. What is given to you is that there is *some* creature with such and such properties, and is thus purely existential. So *objects* are not the content of perceptions; the content of a perception is *made true*, in favorable cases, by an object; but the object is not itself lodged in that content. Objects are the objects of perception only in the attenuated sense in which "Sally met a man" is about Frank or Harry or Doug (whoever it was that Sally met). We must distinguish content from truth-maker.

If we maintain that external objects or even states of affairs *are* constituents of the information borne by your perceptions, then we end up guilty of self-contradiction. As we noted, given only the information that is *in*

the perception, it is possible that you are seeing android-Fido or even that there is no external world at all. So the content of the perception leaves it *open* whether Fido is there – whether anything is there, for that matter. That content, as we said, is Fido-neutral: *that* specifically is what it means to make the true claim, given that perception, it is epistemically possible that Fido isn't there. Given – what nobody denies -- that the perception leaves is epistemically open whether Fido even exists, we must acknowledge that the content of the perception is Fido-neutral – Fido is no part if of it. So if we say that Fido himself *is* part of the content of that perception, then we are forced to say, absurdly, that Fido both is and is *not* part of the information encoded in that perception.

Uncontroversially, given only what the perception tells you, it is epistemically possible that you are seeing android-Fido, or twin-Fido. So if, with the externalist, we say that Fido is a constituent of the content of your perception, then we must say, absurdly, that the content of the perception is not part of what the perception tells you. But what is the content of a perception if not what it tells the percipient? If we separate the content of the perception from what the perception tells you – and this is exactly what the externalist must do – then we end up operating with an incoherent or vacant notion of content.. So we don't want to make that separation. But if we *don't* make that separation, then as we saw, Fido himself ends up *not* being anywhere in that content, in which case externalism is false. So to avoid incoherence, we are forced to say that the content of your perception does not include Fido at all. That content is not *Fido*, nor anything of the form…*Fido*…Rather it is, for reasons that are independently argued, some existence claim that is *made true* by the Fido.

Let us relate this point on Brentano's thesis. When you have a hallucination of a unicorn, the content of your condition is not simply *Pegasus*. It is (say) *there is a horse-like creature with a horn on its head and white fur.*…The *content* of your perception is not a non-existent unicorn. If there *were* a unicorn there, the content of your perception wouldn't be an existent unicorn, as we, in effect, just saw when discussing Fido. The content of your perception is given by an existence claim. What your perception is saying is: *there is a being with such and such properties.*…You don't see a non-existent unicorn. At the same time, your state is obviously *representational*: its representational content is given by an existence claim, and is not identical, even in part, with an unicorn (existing or not existing). So the existence of

hallucinations doesn't threaten BT, and we don't have to ontologize to protect BT.

Once we acknowledge the fundamentally existential nature of sense-perception, it becomes clear that, ironically, externalism is a form of VOP and that internalism is not. The externalist says that the relation between a perception and its content is (epistemically) contingent. According to the externalist, your perception has Fido for at least part of its content. At the same time, everybody agrees that it is *epistemically* possible that you are seeing twin-Fido, and not Fido. In fact, it is epistemically possible that there is *no* dog (or dog-like entity) there at all. In that case, according to the externalist, the *content of your perception doesn't exist* or, at least, it has a gap in it.

The situation is dire. It is *epistemically possible,* as Descartes and Hume pointed out, that there is no external world at all. (I doubt anyone would deny this, even though few would accept the skepticism to which this fact drove Hume and Descartes.) Now the externalist identifies content with something extra-cranial. So if the externalist is right, then it is epistemically possible that your perceptions don't have *any content at all.*

Evans, MacDowell, and other externalists maintain that, where there is no object, there is no content. They are being consistent here – true to their own externalism. *If* externalism is right, then the very *content* of your perception of the dog includes that dog itself. If that is right, then the representational content of a hallucination of a dog has a gap in it where there should be a dog. To the extent that a state is a hallucination, it is *missing* content. Given this, suppose that *all* of our waking life is a dream – this is obviously *epistemically* possible. In that case, the externalist must hold that our experiences are content-free: the (would-be) perceptual states of a brain in a vat or the victim of an Evil Demon become devoid of content.

But this is very hard to believe. In fact, it is self-contradictory. When we say that it is *epistemically* possible that a given perception is a hallucination, we mean that it is possible *given the information that is encoded in that perception.* So talk of what is epistemically possible *presupposes* that perceptions have content. Everybody, externalists included, wishes to admit that it is *an epistemic possibility that*, right now, I'm seeing twin-Fido, as opposed to Fido, or even that I'm having a hallucination and am not seeing anything at all Since everybody countenances the idea that certain things are *epistemically* possible relative to that perception, it follows that everybody, externalists included, are committed to the idea that the perception has a

resilient content – a content that exists in all epistemically possible counter-factual scenarios where that perception exists: scenarios where Fido doesn't exist, scenarios where we are dupes of Evil Demons. So it makes no sense to say that it is epistemically possible that our perceptions have no content. When we say that such and such is epistemically possible with respect to our perceptions, we are saying precisely that the content of our perceptions doesn't rule it out. The externalist says that the content of the perception is extra-cranial. So the externalist is forced to make the false, indeed the absurd, claim that it is *not* epistemically possible that I am seeing twin-Fido (as opposed to Fido) or that I am a hallucinating.

Let us use the term "perceptual state" to describe perceptions and hallucinations alike. (So the word "perceptual" is being used as a mnemonic device, not as a way of denoting the property of *really* being a hallucination.)

The *contentfulness* of our perceptual states cannot be doubted. If we say of some perceptual experience that it *could* be a hallucination, we are saying that *given its content*, it is an open question whether it is a hallucination or not. (This is an open question where every perceptual state is concerned.) So skepticism *presupposes* contentfulness. So if the externalist wishes to be coherent, he must *accept* the contentfulness of our perceptual states – the veridical perceptions and the hallucinations alike. The Evans-MacDowell view is that, in so far as perceptual state is a hallucination, it has a *gap* in it. A consequence of this view is that, if we are victims of Evil Demons, then our perceptual states would turn out to be altogether content-free. But we've just seen that our perceptual states *do* have content, even if it turned out they were all hallucinatory, and that it is incoherent to maintain otherwise. The epistemic possibility of hallucination presupposes the contentfulness of our perceptual experiences. Therefore the Evans-MacDowell view is incoherent, given the very notion of epistemic possibility. So to avoid incoherence, the externalist must actually hold onto the idea that the contentfulness of our perceptual experiences is not jeopardized by the epistemic possibility of global hallucination. But an *externalist* could not do this. For what survives skeptical doubt is precisely what is internal. So in so far as contentfulness is Demon-proof (or vat-proof) it follows that contentfulness is not external, but is inseparable from the perceptual experience itself.

For the externalist, the relation between a perceptual experience and its content becomes epistemically contingent. You can have the perceptual experience: even so, the *content* of the experience might or might not exist.

After all, the *content*, if it exists, will be some external state of affairs – some star that might be long gone or might never have existed to begin with. So the perceptual experience is one thing, its content – what it *tells* the percipient – is another. And the relation between the two is fragile. It is possible, at least epistemically, that *that* very perception might have had a very different content. So for the externalist, the perception has an external relation to its own content: the perception is a veritable intermediary, a veil; its content is on the other side of that veil. When that veil is peeled away, so to speak, the content might be there – or it might not. The root-commitment of externalism thus turns out to be the idea that perceptions are intermediaries: the content of the perception is quite separable from the perception itself. The Humean view, powerfully critiqued by Reid, recrudesces in externalism.

With the internalist the story is different. On the internalist's account, the content of the perceptions is Demon-proof, vat-proof, and generally hallucination-proof. For the internalist, you cannot have the perceptual experience without the content. So there is no layer of perceptions *between* subject and content. Once there is a perceptual experience, there *ipso facto* is the content. In a way, there ceases to be any intervening layer of *representations* at all. There is not, on the one hand, a perception and, on the other hand, a separable content of perception. Where the perception is, the content is as well: the perceptual experience just *is* an awareness of content. So there is not, except nominally, a distinction between the existence of the perceptual experience, on the one hand, and awareness of content. There is just one thing – an awareness of content.

Chapter 24 Empty names

There may or may not be a Yeti. Either way, the sentence "Yeti is ferocious" has *cognitive* content: it *tells* you something. But it is unclear what, if anything, that sentence *literally* means. Suppose there turns out to be no Yeti. In that case, "Yeti" will turn out to be an empty name. So "Yeti", in that case, would be the linguistic counterpart of a hallucination. "Yeti is ferocious" *ought* to encode a proposition of the form *alpha is ferocious.* But if there is no Yeti, then it will actually encode something with a blank in it:__*is ferocious.*

But as we noted, "Yeti is ferocious" seems to *communicate* something more than:__*is ferocious.* It doesn't matter whether there turns out

to be a Yeti or not.

There are several questions here. First, supposing that there is no Yeti, does "Yeti is ferocious" have *any* (complete) proposition for its literal meaning? Supposing that the answer is "yes", what is that proposition? Supposing that the answer is "no", how does "Yeti is ferocious" manage to communicate a significant message (as opposed to something ill-formed like *__is ferocious*)? And what exactly is that message?

In this part of the book, I would like to show how, by combining the semantic and epistemological points made so far, we can make some headway on these matters. So we will give an analysis of *empty names* and, more generally, of *empty singular terms.*

As we will see, there are different kinds of empty names. There are those associated with fiction ("Fred Flintstone"), and there those associated with erroneous beliefs as to the actual world ("Atlantis", "phlogiston"). In this chapter, we will deal with empty names of the latter kind.

It is epistemically possible that the world came into existence a thousand years ago, complete with fake memories and fake evidence of times past. (Russell made this point.) If that is the case, then "Socrates" is empty – it refers to nothing. And "Socrates was wise" doesn't encode any proposition.

But "Socrates was wise" obviously has *cognitive* significance. No matter when the world came into existence, that expression has cognitive significance.

One minute ago, you met Fred. Fred has never heard the word "Socrates" before, but is otherwise perfectly normal. You say to Fred: "throughout history, there has been only one philosopher to drink Hemlock. 'Socrates' was his name."

In fact, we can do better than this. Suppose that Fred is a very sophisticated philosopher of language; he just happens, for some strange reason, never to have heard the name "Socrates" (or any translation thereof). You say to Fred:

> "somebody x was uniquely a great philosopher of antiquity who died of Hemlock poisoning. 'Socrates' is a non-connotative singular term referring to Socrates."

Under the circumstances described, "Socrates" doesn't refer to anything. In those circumstances, the world is only a thousand years old. But

(SW) "Socrates was wise"

will have plenty of cognitive significance for Fred. For the reasons just discussed, that cognitive significance will be:

(HK) *somebody x was uniquely a great philosopher who drank hemlock, and x was wise.*

For any predicate *…x…*, an exactly similar argument shows that, for Fred, the cognitive significance of

(S) "…Socrates…"

is

(F) "*somebody x was uniquely a great philosopher to drink hemlock and…x…*"

So sentences of the form "…Socrates…" will be replete with cognitive significance for Fred even though "Socrates" is an empty name.

There is more to say. Since (S) communicates (F) to Fred, it follows that:

(HKP) "In some possible worlds, Socrates did not drink hemlock".

will be ambiguous, in terms of what it communicates (not in terms of what it literally means) between:

(HPKW) *in some possible worlds, somebody x was uniquely a great philosopher who drank hemlock, and x was wise, and x did not drink hemlock*

and

(HPKN) *somebody x was (as a matter of fact) uniquely a great philosopher who drank hemlock, and x was wise, and in some possible*

worlds: x did not drink hemlock

In terms of what it *literally* means, (HKP) is not ambiguous at all. It unambiguously means:

(HPKP) In some possible worlds, Socrates did not drink hemlock.

But, for reasons we've discussed, (HPK) will always *communicate* some existence-claim. And, for reasons we've discussed, we know that, where Fred is concerned, that existence claim will be *either* (HPKW) or (HPKN).

Fred knows that "Socrates" is a singular term. Fred will know that the person in fact referred to by "Socrates" *might* not have become a philosopher and *might* not have drunk hemlock. Fred will know that "Socrates" refers to the guy who *in fact* was a great hemlock-drinking philosopher. But Fred will also know that *that guy* might never have become a philosopher. After all, Fred knows that any person's life *might* have gone differently.

All of this is perfectly consistent with the cognitive significance of (SW) being (HK). Suppose I say to you. "Exactly one person was a great philosopher who drank hemlock. But in some possible worlds *that* very guy did not drink hemlock." That is a perfectly reasonable claim. (SW) communicates (HK) to Fred. But because Fred speaks English, he will know that "Socrates" is a term that picks out some specific individual. Because he knows this, if Fred hears (HPK), he will give the modal-operator *narrow*-scope: he will take it to mean (HPKN), not (HPKW). So Fred will rightly *not* take (HPK) to encode the contradictory claim that, in some world, somebody did and did not drink Hemlock; he will rightly take to encode a contingent proposition; Fred will rightly assign the right modal status to (HPK).

Dummett's wide-scope descriptivism

There is a *prima facie* problem for my analysis. This problem can best be understood in terms of a thesis of Michael Dummett's.

In Dummett's[144] view, "Socrates" *is* identical with a definite description. But when "Socrates" occurs within the scope of certain kinds of

operators (especially, modal operators), it is to be treated as a Russellian-quantifier that is given wide-scope with respect to that operator. So, in Dummett's view, the literal meaning of (HPK) is (HKPN). Thus, for Dummett, "Socrates" is a "wide-scope" definite description.

Obviously my position resembles Dummett's. This is not good; for it is pretty clear that Dummett's analysis is wrong.[145] "Socrates" is *not*, at the level of literal meaning, a description of any kind: not a wide-scope or a narrow-scope description.

For the sake of argument, suppose that "Socrates" was a wide-scope description for its meaning. In that case, the following would be analytic:

(SNW) "Socrates was wise. Therefore somebody drank hemlock."

For (SNW) would be synonymous with:

(SNW*) "Somebody x was (as a matter of fact) a uniquely a great philosopher who drank hemlock, and x was wise. Therefore somebody drank hemlock."

But those two sentences are obviously not synonymous. Suppose you learn the meaning of "Socrates" thus. Somebody says to you:

"Plato wrote many dialogues. Exactly one person x was a protagonist in most of the dialogues. "Socrates" refers to x."

As it happens, you never learn that x (so to speak) drank hemlock. (This is a perfectly possibly scenario. Surely there are people who know what "Socrates" means who never learn that Socrates died of hemlock-poisoning.)

Under that circumstance, you will know the semantics of "Socrates" as well as anyone. But you will rightly *reject* (SNW) as being a non-sequitur. So Dummett is wrong; names are not wide-scope descriptions.

But my position is not Dummett's. Dummett is talking about literal meaning; I am talking about *communicated* meaning. Here, as before, we must be extremely careful about where we put the relevant operators. According to Dummett, the correct analysis of "Socrates" is:

The semantic rule for "Socrates" is: "…Socrates…" means: exactly

one great philosopher x drank hemlock and…x…,

with the qualification that, when "Socrates" occurs in the scope of certain operators, the just mentioned quantifier is to be given wide-scope with respect to them.

For me the right analysis of "Socrates" is simply:

The semantic rule for "Socrates" is this: "Socrates" refers to Socrates, i.e. the "…Socrates…" literally means:…Socrates….

So if I say to you

"Socrates" is the great philosopher who drank hemlock,

what I am saying is:

Somebody x was a great philosopher who drank hemlock, and the semantic rule for "Socrates" is: "Socrates" refers to x, i.e. "… Socrates…" means:…x…

So, on my view, even when descriptions are being used to *fix the referent* of "Socrates", I am, within the context of that description, giving *narrow* scope to the operator "the semantic rule for ‘Socrates’ is". By contrast, Dummett would give it *wide*-scope. Dummett’s position is (this is not a quotation):

The semantic rule for

"…Socrates…"

is:

somebody x was uniquely a great philosopher who drank hemlock and…x…,

with the qualification that in modal contexts, the quantifier is to be

*given wide scope***.**

My analysis is:

> *Somebody x was uniquely a great philosopher who drank hemlock, and the semantic rule for "Socrates" is "…Socrates…" means: …x…*

So, on my view, ‘Socrates" is not a description at all: not a wide-scope or a narrow-scope description.

Nonetheless, there might *still* seem to be a problem with my analysis. For on my view, the *communicated* meaning of "…Socrates…" is some existence claim. Of course, it communicates entirely different existence claims to different people, depending on *how* they were first told the meaning of "Socrates". But it always communicates *some* such claim. And this, it might seem, makes my analysis coincide with Dummett’s.

This is not so. Anyone who speaks English knows that "Socrates" is a label. So whenever one hears a sentence like "Socrates was the great philosopher who drank hemlock", one *in effect* (tacitly) gives the operators involved the appropriate amount of scope. Suppose Joe asks "who was Socrates?", and you say: "Socrates was the great philosopher who drank hemlock". Joe knows that what is being said is:

> *Somebody x was a great philosopher who drank hemlock. And the semantic rule for "Socrates" is: "Socrates" refers to x.*

Joe knows that what is being said is *not*:

> *The semantic rule for "Socrates" is: Somebody x was a great philosopher who drank hemlock, and "Socrates" refers to x.*

Joe will correctly give "wide-scope" to the operator "the semantic rule for__is" operator. Let us deal with an objection to this line of thought:

You say repeatedly that everybody who speaks English knows exactly what degree of scope to give to these operators. You say that the average Joe knows with complete clarity that, when he is told

(*) " 'Socrates' refers to the great philosopher who drank hemlock"

what is meant is:

Somebody x was a unique great philosopher who drank hemlock, and the semantic rule for "Socrates" is: "…Socrates…" means:…x…

So, on your analysis, Joe knows with complete clarity that, when he is told (*), he is not being told that:

"The semantic rule for "Socrates" is: "…Socrates…" means: exactly one person x was a great philosopher who drank hemlock and…x…

But you are imputing far too much semantic knowledge to people. Russell surely had more semantic knowledge than practically anyone who has ever lived. And yet, as Kripke showed, he was wrong about the semantic rule for "Socrates".

There is an obvious reason why this objection fails. Linguistic competence consists in *tacit* or *sub-personal* knowledge. Any four year old can understand enormously complex sentences, and there are extremely compelling reasons to suppose that this involves knowledge of grammatical, transformational, phonological, syntactical, and even pragmatic rules. Obviously the four year old doesn't have discursive, conscious knowledge of such rules. He has *tacit*, sub-personal knowledge of them. When I say that the average Joe "knows" such and such about semantics, I am obviously talking about tacit knowledge. When we say that Russell or Chomsky have more semantic knowledge than other people, we are saying that they have more *conscious*, discursive knowledge. In the tacit, sub-personal sense, Russell obviously knew the semantics of "Socrates". It was Russell's discursive, conscious view that was wrong. So the objection just tendered has no force. The only way to give it force would be to insist (as, shockingly, some

philosophers have done) that all semantic knowledge is possessed *consciously.* But even this would fail to give the objection force. For we could then reconstruct our counter-objection by distinguishing conscious semantic knowledge from conscious semantic *meta*-knowledge. The four year old thus has conscious semantic *knowledge* of what is meant by "the man in the corner is a professor". What the professional semanticist has that the four year old lacks is knowledge *about* that knowledge: meta-knowledge – knowledge of what it is that is known. So even if we take the Herculean approach of denying that there is sub-personal knowledge – not an approach I condone – the objector's point is *still* innocuous.

Chapter 25 Fictitious names

Sentence-tokens of the form:

"…Sherlock Holmes…"

are replete with cognitive value. The sentence:

(*) "For the first time in his life, Sherlock Holmes didn't know what to do"

is *cognitively* significant. But there is no Sherlock Holmes. So "Sherlock Holmes" doesn't refer to anybody; it is a blank. So presumably (*) is synonymous with the non-sentence:

(**)"For the first time in his life, __ didn't know what to do"

How are we to explain the cognitive significance of the aforementioned sentences?

We've already explained the cognitive significance of empty *non*-fictitious names. We've already explained how, if "Socrates" turned out to be empty,

"Socrates was wise"

could have cognitive value. Perhaps, it might be thought, we can simply apply what we said in that context to fictitious names.

For reasons Kripke discusses in his Locke lectures, this is not the case. Empty fictitious names must be dealt with differently from empty non-fictitious names.

Russell had a very good explanation as to why "Sherlock Holmes was speechless" has cognitive value. In this view,

"…Sherlock Holmes…"

is synonymous with some existence claim:

"somebody x was uniquely a pipe-smoking detective […] and…x…"

There are a couple of problems with this. Consider the sentence:

(1) "Socrates smoked a pipe"

We know that there is some object O such that, at the level of semantics, the meaning of (1) is:

(2) *O smoked a pipe.*

Obviously

(3) "Sherlock Holmes smoked a pipe"

has the same semantic structure as (1). But suppose (3) meant:

(4) Something x was uniquely a clever detective who […] and x smoked a pipe.

In that case, (1) and (3) have radically different semantic structures. But this simply isn't plausible.

There is another reason to reject the Russellian approach; this one corresponds to Kripke's criticism of Russell's analysis of proper names.

Let *…x…* be any *non-trivial* predicate. By that I mean a predicate that can informatively be predicated of a term denoting a spatio-temporal object. If I tell you that Bob is either a square or not a square, or that he is extended in space, I haven't really told you anything about Bob. If I tell you that Bob smokes a pipe, I have told you something substantive about him. So "not both a square and a circle" is trivial; so (in this context) is "extended in space". But "red", "smart", "smokes a pipe" and the like are non-trivial.

For any non-trivial predicate *…x…*, the sentence:

….Sherlock Holmes…

can be informative to somebody who *does* know the semantics of "Sherlock Holmes". But if "Sherlock Holmes" were a description quantifier, this would not be the case. For at least *some* non-trivial predicate *…x…*, the sentence:

…Sherlock Holmes…

would encode a tautologous proposition. Suppose

"…Sherlock Holmes…"

meant:

exactly one thing x was a detective who played the violin and had a sidekick named ‘Watson’ and…x…

In that case,

(i) "Sherlock Holmes had a side-kick named ‘Watson’"

would communicate a tautology to anyone who knew the semantics of "Sherlock Holmes"; for (i) would then mean:

(ii) exactly one thing x was a detective who played the violin and had a sidekick named ‘Watson’ and x had a sidekick named "Watson".

But we can imagine somebody who knows the *meaning* of "Sherlock Holmes" (so far as it has a meaning: see below) to whom (i) communicates a non-tautologous proposition. Suppose the only Sherlock Holmes story Larry has read is the one in which Watson isn’t mentioned. In that case, Larry will understand (i) perfectly well, *and* it will be informative to him.

Even if, by chance, Watson *is* mentioned in every Sherlock Holmes story, this wouldn’t affect the cogency of the point just made. Imagine a counter-factual scenario in which the following is the case. Conan Doyle wrote a few more Sherlock Holmes stories than he actually did. In those stories, Watson isn’t mentioned; neither is the act that Holmes played the violin. Larry happens to read just those stories. Larry has no idea that Sherlock Holmes ever had a sidekick or that played the violin.

Given *any* expression C denoting fictional character, and given *any* non-trivial predicate *…x…*, you could always identify a scenario – it might or might not be counterfactual -- in which *…C…* is non-trivial to somebody who was acquainted with the semantics of C. So the Russellian analysis is not viable.

So we are left with no solution to the question: How can sentences of the form "…Sherlock Holmes…" be cognitively significant?

There is an obvious answer: we say about "Sherlock Holmes" what we said earlier about empty non-fictitious names. A person is always *introduced* to the meaning of a name through descriptive information. That information is given wide-scope in the definition, so that the expression in question is a mere label, not a quantifier. But, inevitably, that descriptive information is implicated in the *cognitive* significance that sentences involving that expression have for one. One is introduced to the expression "Sherlock Holmes" through some passage like:

"There was a clever detective who lived on Baker street. He played the violin and was addicted to opium. His name was Sherlock Holmes…"

Thus, for reasons we've already considered, the cognitive significance of

(3) "Sherlock Holmes smoked a pipe"

would be

(4) Something x was uniquely a clever detective who lived on Baker Street […] and x smoked a pipe

But (4) would not be the *literal* meaning of (3); it would only be its cognitive meaning.

This proposal initially seems reasonable enough. But it is not *quite* right.

As we've noted, reference is always *fixed* through a wide-scope description. The meaning of "Socrates" is given through an existence claim:

There is some x such that x was a great philosopher of antiquity who drank hemlock and "Socrates" refers to x.

(Recall that, even when "Socrates" is being defined ostensively, it is really being defined through a wide-scope description.)

Our analysis of "Sherlock Holmes" amounts to this: one learns its semantics through a wide-scope existence-claim, some claim along the lines of:

> *(SH) Something x was uniquely a clever detective who lived on Baker Street […] and "Sherlock Holmes" refers to x.*

For now familiar reasons, the wide-scope description may figure in the cognitive content of sentences like "Sherlock Holmes smoked a pipe". But it will not figure in the literal meaning of such sentences.

All of this seems perfectly fine; it seems as though everything we said about empty non-fictitious names can be mapped onto "Sherlock Holmes".

But there is a problem. This problem was identified by Kripke[146]. Suppose that, by sheer coincidence, there actually *was* a unique detective x such that x lived on Baker Street, and so on, and x smoked a pipe. Let O be that detective. Under that circumstance, would "Sherlock Holmes" refer to O? It would not. Given any *actual* object, no matter what that object has done, or might have done, "Sherlock Holmes" will not refer to that object. Even if there was an actual detective who, unbeknownst to Conan Doyle, did all of the deeds attributed by Doyle to Holmes, "Sherlock Holmes' would *still* not refer to that person. In fact, even if that person were named "Sherlock Holmes", nonetheless occurrences of "Sherlock Holmes" in Conan Doyle's stories *still* wouldn't refer to that person. They would be homonyms; but they wouldn't co-refer. In fact, even if Conan Doyle *knew* of that person, and *knew* that he was named "Sherlock Holmes", and *knew* that he performed those deeds, occurrences of "Sherlock Holmes" in his stories *still* wouldn't refer to that person. So long as Doyle was writing *fiction*, "Sherlock Holmes" would not refer to anyone actual. It would make no difference if the fiction were horribly unoriginal and its content coincided with occurrences in the

real world. So long as Conan Doyle was writing with the intention of producing *stories*, and not journalism or historiography, occurrences of "Sherlock Holmes" in those stories would not refer to any actual person, no matter what that person was called and no matter what that person did. Any correct analysis of the semantics and pragmatics of "Sherlock Holmes" must be consistent with this fact.

According to the analysis we put forth a moment ago, the meaning of "Sherlock Holmes" is given through (SH) or, at any rate, some similar existence claim. But if that analysis is right, then "Sherlock Holmes" would refer to an *actual* person x such that x uniquely was a clever detective who lived on Baker street, played the violin, and so forth. So if that analysis is right, then the occurrences of "Sherlock Holmes" could, potentially, refer to an *actual* human being: if somebody had the relevant properties, the occurrences of "Sherlock Holmes" in Doyle's stories would refer to him.

But this is not the case. For reasons we've just seen, the occurrence "Sherlock Holmes" couldn't possibly refer to *any* actual person. So the analysis in question is false. So we are *still* stuck with no answer to the question: Why are sentences of the form

"…Sherlock Holmes…"

cognitively significant?

The supposition-operator

But an answer is forthcoming; we need only modify slightly the analysis that we just rejected.

When somebody tells you a story, there is an understanding that what they are saying is not be taken as fact. So, in effect, the entire story is prefixed by a kind of disclaimer:

The following is not to be taken as fact.

This disclaimer can be thought of as an operator. The standard way of expressing such an operator is to use words like "suppose that" or "let it be

the case that". If I am a physicist, I will want to talk about physical objects in general, not about this or that specific object. I will say:

"let x, y, and z be any objects whose masses and states of motion are…"

This is equivalent to:

"Suppose that x, y, and z are objects whose masses and states of motion are…"

Suppose I say: "suppose there is some man x who is a great detective…" I can use expressions to refer to the man who, within the scope of the supposition, is described. I can use demonstratives and definite descriptions; I can say "he had a terrible secret" or "the great detective we spoke of earlier wanted to buy a house in Florida". We can also give x a name; we can call him "Sherlock Holmes" or "Fred Flintstone". So we can use all the referential weaponry afforded us by the language we are using.

But these referring terms will never reach beyond the "suppose that" operator; they are blocked by it; they will refer only *within* the scope of the supposition.

Consider the following passage:

Suppose the following. A guy is walking down the street. He wants to buy a new car. His name is "Chip Wolcott". Chip hates his boss, whose name is "Doug". Chip is married to Doug's daughter. But, for some reason, this makes Doug behave in an even more autocratic and contemptuous way towards Chip than he would otherwise. Chip's wife, Betsy, idolizes her father, and will not tolerate even a hint of ill-will towards him. Betsy is the only person in the world Chip can speak at all honestly with. But the only thing Chip wants to talk about is how much he hates Doug…

Imagine that there is, in actuality, some person O, such that O's

name is "Chip", and O has a boss named "Doug", and so on. Do the occurrences of "Chip" in that passage refer to O? No. The first three words of the passage prevent that.

Of course, a person can use a story, or a hypothetical claim, to *convey* a true statement. A lawyer trying to strike a deal with the prosecutor might say: "hypothetically speaking, suppose I had a client who knew where the body was buried…" The lawyer is making it pretty clear that his client *does* know where the body is buried. But that is not what he is *saying*. Rather, it is what is being communicated – it is a matter of innuendo or implicature. From a strictly semantic point of view, the words "hypothetically speaking, suppose that…" neutralize everything said thereafter. *Technically* none of it incriminates the client; for *technically* nothing has been admitted. The words "suppose that" strip everything thereafter of truth or falsity.

This is not the place to say exactly what suppositions are. In this context, it is enough to register the fact that there are such things.

Every piece of fiction is, in effect, prefixed by the "suppose that" operator. So given a passage beginning with:

> *There was a great detective who lived on Baker Street. His name was "Sherlock Holmes". He played the violin….*

the existence claim (*for some x, x was a great detective […] and "Sherlock Holmes" refers to x*) falls within the scope of the "suppose that" operator; consequently, it is neither true nor false. After all, what is being said is not: *there was a great detective who lived on Baker Street.*…Rather, it is:

> (*) *suppose that there was a great detective who lived on Baker Street […] and "Sherlock Holmes" was his name…*

(*) is not true or false. Even if there is a great detective who lives on Baker Street, and is named "Sherlock Holmes", the occurrence of "Sherlock Holmes" won't refer to him. The reason is that *no* statement of the form "suppose such and such" is true or false. Even if "S" is true, "suppose S" is neither true nor false. Even if

> (*) "there was somebody x who lived on Baker Street […]"

is true, the sentence that results when (*) is pre-fixed by a "suppose that" is neither true nor false. Every claim occurring in a story is, in effect, prefixed by such an operator. Claims like "his name is ‘Sherlock Holmes’" are no exception.

Now we can explain why "…Sherlock Holmes…" is cognitively significant, even though "Sherlock Holmes" doesn’t refer. Every story can be thought of as being given by a claim that begins with a "suppose that". Within the scope of that operator, the semantics of "Sherlock Holmes" is fixed by a wide-scope definite description. For reasons with which we are familiar, that descriptive information will figure in the cognitive significance of "…Sherlock Holmes…", but not in what is semantically encoded in them. In fact, *even within the scope of the "suppose that" operator,* that descriptive information will not figure in the *semantics* of "…Sherlock Holmes…" What is being supposed is that "Sherlock Holmes" is a name, like "Socrates". So, within the scope of that supposition, "Sherlock Holmes" is no more a quantifier than "Socrates"; and, within the scope of that supposition, descriptive information is used to fix the referent, but not give the meaning, of "Sherlock Holmes". So, within the scope of that operator, "Sherlock Holmes" is a proper name.

Why, strictly speaking, "Sherlock Holmes" isn’t even an empty name

But this discussion has an interesting consequence. "Sherlock Holmes" is a name *only within the scope of a supposition* .Of course, some actual person might be named "Sherlock Holmes". But *that* name would be a mere homonym of the "Sherlock Holmes" that occurs in fiction.

So *in English* – I mean, in the actual language that we speak – not only does "Sherlock Holmes" not denote anything: it isn’t even *supposed* to mean anything. It is *supposed* to be supposed to denote something; but it isn’t supposed to denote anything. Let me explain.

Imagine that, with the agreement of all other English-speakers, I define an expression "Argo" as follows:

(AR) If somebody x uniquely smoked a cigarette yesterday, then

"…Argo…" means:

…x…

If there is no such person, then

"…Argo…"

is abortive.

"Argo" is an empty referring term. (Strictly speaking, tokens of "Argo" are empty. The *type* "Argo" has a function for its meaning: functions from contexts to referents. So the type is not empty. To simplify discussion, let us henceforth ignore the type-token distinction; it has no relevance in this context) But, empty though it is, "Argo" is now a part of our language. (AR) is a true statement *about* English (or, at any rate, about an extension of English). By contrast, the following is *not* a true statement about any language:

(SH) If there is a unique x such that x was a great detective who lived on Baker Street […], then

"…Sherlock Holmes…"

means:

…x…

If there is no such person, then

"…Sherlock Holmes…"

is abortive.

The English language does not comprise (SH). No language does. What exists is a *supposition* to the effect that (SH) exists. What exists is a *supposition* to the effect that there was a detective x who had a sidekick

named Watson and x smoked a pipe and that "Sherlock Holmes" names x […] But (SH) is not a semantic rule of English.

Here we must be careful. When we say things like "Sherlock Holmes smoked a pipe", our words are prefixed by an implicit "suppose that" operator; in fact, they are prefixed by much more than that. When one talks about "Sherlock Holmes" or "Fred Flintstone", it is understood that one is operating within the scope of a supposition that was initiated long ago, and that has been continued by various movies, cartoons, books, and so on. If one were to say "Sherlock Holmes smoked a pipe", but were able to say that *outside* the confines of that supposition, the result would be gibberish; it wouldn't even be in the same category as "Argo smoked a pipe". Of course, there might be a sentence-token *homonymous* with "Sherlock Holmes smoked a pipe" that was not gibberish. But that is irrelevant. If somebody managed to use the term "Sherlock Holmes" in such a way that it was not a mere *homonym* of the term that occurs in Doyle's stories, and in such a way that it was not internal to the kind of supposition of which we've been speaking, the result would be absolute nonsense. Suppose a child says "Sherlock Holmes was obviously a very determined man", thinking that the story he was reading was a piece of history. In this case, he is not operating within the scope of the supposition we've been discussing; *and* his use of "Sherlock Holmes" is not a mere homonym of Conan Doyle's use. Under that circumstance, the child's words would be nonsense; they wouldn't even make it into the same category as "Argo is bald". "Argo is bald" contains a constituent that is *supposed* to refer, but doesn't. As a consequence, it encodes no proposition. The child's words comprise a constituent that isn't even *supposed* to refer: "Sherlock Holmes" isn't supposed to refer; it is *supposed* to *suppose* to refer. So the child's words are even *more* meaningless than "Argo is bald". We've already explained why, this fact notwithstanding, the child's words would have cognitive significance for his auditors and also for himself.

Empty non-fictitious terms versus empty fictitious terms

We must distinguish *empty* singular terms from *fictitious* singular terms. Astronomers once posited the existence of a planet between Mercury and Venus. They called it "Vulcan". Vulcan turned out not to exist.

The term "Vulcan" was supposed to refer. It is not a term of fiction. By contrast, "Sherlock Holmes" and "Fred Flintstone" are not even supposed

to refer.

Sometimes we speak of so-called "mythical" objects: Zeus, Hera, and the like. Mythical objects are not fictitious objects. "Zeus" – or its ancient Greek translation, rather – was supposed to refer. The gods were not *fictions*; they were posits of erroneous *beliefs*. (Of course, where there is wishful thinking, fiction and erroneous belief may coalesce. But let us idealize away from this psychoanalytic side-light.) Zeus and Hera are in the same category as Vulcan and phlogiston. They are not in the same category as Fred Flintstone and Sherlock Holmes.

Of course, *we* now take it for granted that "Zeus", "Hera", and the like, are empty. So, to us, they are comparable to "Fred Flintstone" and "Sherlock Holmes". When *we* use these terms, *we* are creating fiction. But to an ancient Greek, those expressions (or their translations) were not terms of fiction.

Many direct reference-theorists have posited the *actual* existence of Zeus and Fred Flintstone. They have posited the *actual* existence of mythical and fictitious objects. It is clear why they have done so. A corollary of direct reference theory is that a singular term has *no* literal meaning if it doesn't refer. So, strictly speaking, "Fred Flintstone is vulgar" has *no* literal meaning unless "Fred Flintstone" refers. That sentence obviously has cognitive significance. This last fact seems to be inconsistent with its having *no* literal meaning. So, it is thought, we must suppose that "Fred Flintstone" *does* refer.

This reasoning is entirely spurious. Of course, direct reference theory *does* have to deal with the fact that sentences containing empty singular terms may have cognitive significance. But this is no problem at all. We need to distinguish literal meaning from communicated meaning. And we need to register a few fairly obvious facts about epistemology – in particular, that it is states of affairs involving objects, and not just objects *simpliciter*, that are perceived. Given these pedestrian facts, it becomes clear why "… Fred Flintstone…" is cognitively significant, even though it contains an empty singular term (even though, to put it more accurately, it contains a term that is *supposed* to be *supposed* to refer, but doesn't). Give or take some niceties relating to the "suppose that" operator, exactly similar remarks apply to "…Vulcan…", "…phlogiston…", and "…Sherlock Holmes…".

Most direct reference theorists do not see things this way. They have simply taken it for granted that "Fred Flintstone" and "Zeus" *must* refer, given that sentences containing them are, in some sense, significant.

There is an obvious problem with this move. Opponents of direct reference theory will see it as proof that it is false: if direct reference theory requires the *actual* existence of things which we know not to exist – Fred Flintstone, Zeus, Vulcan – then direct reference theory is wrong. It is as wrong as the theories that posited phlogiston and Vulcan.

Also, it is obviously silly to posit the existence of Fred Flintstone. Direct reference theory ends up being a comedy routine; its advocates are reduced to sniveling children who cling to their toys. Aside from Kermit the frog and the Cookie monster, everybody comes out a loser.

Given this, direct reference theorists have tried to produce *independent* proof that mythical and fictitious objects exist. The most cogent of these is due to Nathan Salmon.

Any sentence containing a singular term permits of an existential generalization.

(i) "Smith is a man",

implies

(ii) for some x, x is a man.

Given a *molecular* sentence containing a singular term, *two* existential generalizations are associated with it; this is because the existential quantifier can be given different degrees of scope. Depending on how it is read,

(*) "John believes that Smith is a man".

entails either

($*_W$) There is some x such that John believes: that x is a man,

or

($*_N$) John believes: that there is some x such that x is a man.

The second says that John believes some existence claim; the first

says that he believes some singular proposition.

Thus, depending on how it is read,

(z) "John believes that: Zeus is a god"

entails either

(z_W) There is some x such that John believes: that x is a god,

or

(z_N) John believes: that there is some x such that x is a god,

Given this, consider the sentence:

(Z) "Fred and Bob both believe that Zeus is a benevolent god."

Everybody will agree that, on some legitimate reading, (Z) is true or, at least, that it *could* be true. (I myself think that (Z) is neither true nor false *unless* "Zeus" refers. So I won't concede this joint in Salmon's argument and, I would argue, in assuming that (Z) is either true or false, Salmon is begging the question at issue. But right now I am simply stating Salmon's argument, not appraising it.) One doesn't have to accept any particular semantic or theological view to grant that (Z) is true. Atheists and polytheists will agree that (Z) can be true; descriptivists and direct reference theorists will also agree on this. So if we suppose that (Z) is true, we are not begging any questions; we are making a supposition that is neutral in every respect: theologically and semantically.

We know that (Z) licenses an existence claim of some kind. What is that claim? The obvious candidate is:

(Z_W) There is some x such that both Fred and Bob agree that x is a benevolent god.

Of course, if (Z_W) is the right existence claim, then it follows immediately that "Zeus" refers to some *actual* entity. So if we are to *deny* that "Zeus" refers – if we are to *deny* the existence of Zeus – then we must generate a *narrow*-scope existence claim that is consistent with (Z). But this, says Salmon, cannot be done. Consider:

(Z_N) Fred and Bob agree that there is some x such that x is a benevolent god.

According to Salmon, (Z_N) is *not* what is entailed by (Z). (Z_N) is too weak. Fred and Bob don't just agree that there is *some* benevolent god. They agree on *which* specific god is benevolent. The only way to capture this fact is by giving wide-scope to the existential quantifier. (Z_W) is the right inference, not (Z_N).

In general, when supposedly empty singular terms fall within the scope of epistemic operators, we must countenance the *wide*-scope generalization. Otherwise we are not being true to the content of the original sentence.

There are many very obvious problems with this argument. What is at issue is precisely whether "Zeus" refers. If "Zeus" does *not* refer, then

(B) "Zeus is a benevolent god"

won't have *any* literal meaning. If (B) is without literal meaning, then it is neither true nor false. If (B) is neither true nor false, then (Z) is not true or false.

Even if "Zeus" is empty, (B) will be replete with cognitive significance, and that cognitive significance will be an existence claim. Let E be that claim. Thus, while having no literal meaning, (Z) will be cognitively significant. Its cognitive significance will be some proposition whose ingredients are E and the epistemic operator "John and Bob agree that". The epistemic operator can be given either wide-scope or narrow-scope. Salmon replies thus:

If you give the existence operator wide-scope, then I've made my case. If you give it narrow-scope, then the resulting claim won't do justice to (Z); it will be too weak. You see, what Bob and John agree on is not just that some god or other is benevolent; they are in agreement as to the benevolence of some specific god.

This response begs the question. Salmon insists that they are in agreement as to the benevolence of some *specific* god. This is just another way of saying: there is some x such that both Bob and John agree that x is benevolent.

What I would say is: there is *no* x such that Bob and John agree that x is a benevolent god. Bob believes that some existence claim holds, some claim like: *there is a bearded god x who lives on Mt. Olympus and x is benevolent.* John believes that same existence claim to be true. When we say that Bob and John agree that *Zeus* specifically is benevolent, that is an imprecise and brief way of saying that they both accept that existence claim. We are tempted to say:

(S) "Bob and John are in agreement as to the benevolence of some specific god".

But (S) is exactly what is at issue. (S) is just a devious way of saying:

(S) there is some x such that Bob and John are in agreement that x is a benevolent god.*

(S*) is exactly the proposition that Salmon wants to prove. But he uses it as a *premise* in his argument. This is a classic case of begging the question.

As somebody who believes that "Zeus" does not refer, I would say: There is *no* x such that John and Bob both agree that x is a benevolent god.

When we say things like (S), we are sloppily expressing the true proposition: there is some specific *existence* claim that both John and Bob believe.

Suppose that little Timmy and little Davie both believe in the Easter Bunny. They both believe that the Easter Bunny has big ears. So

(BE) "little Timmy and little Davie both believe that the Easter Bunny has big ears"

has a true proposition for its cognitive significance. Given this, we might say:

(*) "Little Timmy and little Davie don't agree merely that *something* or other has big ears. They both believe that some *specific* entity has big ears."

Like (BE), (*) has a true proposition for its cognitive significance. But, in so far as it communicates something true, what (*) means is that there is some *specific* existence claim that both little Timmy and little Davie accept. That claim is something like:

(E) there is a big rabbit who hides multicolored eggs in various places on Easter day.

What makes (BE) true is *not* that, for some x, both little Timmy and little Davie believe some proposition of the form:…*x*…What makes (BE) true is that both accept some particular existence claim. For obvious reasons, an expeditious way of communicating that claim is to say:

(BE) "little Timmy and little Davie both believe that the Easter Bunny has big ears".

But no ontological stock is to be put in this fact. As we've seen, sentences containing singular terms *always* communicate existence claims. It is no surprise that they can be *used* to communicate existence claims. In any case, Salmon's argument is a simple case of begging the question.

But it is all a tempest in a teapot; there is no need for any ontologizing. Given a few stand-byes of pragmatics and epistemology, the cognitive significance of "…Zeus…" is readily accounted for without positing the existence of any strange gods and without resorting to the view that "Zeus" is a quantifier.

Empty descriptions

Sentence-tokens like "the 9ft man is kind" are *cognitively* significant, even though they don't semantically encode complete propositions. Since "the 9ft man" is empty, such sentence-tokens mean:__*is kind.*

Of course, one can side-step this by saying that "the 9ft man is a quantifier", not a singular term. But the basic problem will still remain. Let us stipulate that "ze 9ft man" be a singular term that refers to unique person x who is 9ft tall, and that doesn't refer if there is no such person. A token of "ze 9ft man is kind" will have plenty of cognitive significance, even though, by our stipulation, it has no complete proposition for its literal meaning.

Salmon has argued apparently empty singular terms – "the [ze] 9ft man", "the even prime greater than two", "the square circle" – *do* refer. His arguments for this are identical – numerically identical – with his arguments for the existence of Vulcan and Zeus. We've already seen why those arguments fail.

We've also seen why "the [ze] 9ft man is kind" is cognitively significant, even though it contains an empty singular term. The rule for "the 9ft man" is this:

If there is an x such that x is uniquely a 9ft man, then a token of

"…the 9ft man…"

will encode the proposition:

…O…

If there is no unique 9ft man, then

"…the 9ft man…"

will encode no proposition.

It follows, for reasons discussed earlier, that "…the 9ft man…" will *communicate* (though it won't semantically encode) a Russellian existence claim. There is no need to posit the *actual* existence of such a man to account for the cognitive value of such tokens. Exactly similar considerations show why, without resorting to Russell's theory of descriptions, we *don't* have to posit the actual existence of round-circles to account for the cognitive value of "the round circle is made of gold".

Chapter 26 Anaphoric reference and the referential-attributive distinction

The main argument for Russell's theory is that it explains the cognitive difference between

(i) "the inventor of bifocals snored"

and

(ii) "the first post-master general snored".

We've seen that the referentialist position explains this cognitive difference as well as Russell's, *and* fits better with other data than Russell's

theory.

But there are other arguments for Russell's theory. The best argument for it is given, not by Russell, but by Gareth Evans. Ironically, the key to refuting these arguments lies in some ground-breaking points made by Evans himself: points relating to "anaphoric" reference.

Consider the sentence:

(DA) "The father of each girl at the party loved her very much."

Here "the father" doesn't refer to any specific person; it doesn't refer to Jones or Smith or Brown. What (DA) seems to mean is:

(DA_R) For any x, if x is a girl at the party, then for any y such that y is a father of x, y loves x very much.

So "the father of each girl" seems to be a quantifier. This seems to support Russell's view that definite descriptions are quantifiers.

It is easy to generate sentences in which definite descriptions function not as singular terms, but as quantifiers:

(SM) "If a strong man and a weak man arm wrestle, the strong man will win."

Here "the strong man" seems not to refer to anybody. It doesn't refer to Bob or Tim or Fred.

I grant that in (DA) and (SM), the definite descriptions don't refer to anybody. I will argue that this neither falsifies the referentialist analysis nor verifies the Russellian view. This is going to take some work to show. We must start with some points that, on the face of it, don't have anything to do with definite descriptions.

Consider the sentence

(L) "If a man smokes, he damages his lungs".

In (L), the word "he" is functioning *anaphorically.* In connection with this, it doesn't refer to any *actual* person; it doesn't refer to Bob or Fred or Larry. This seems to suggest that "he" isn't always a singular term. I wish to show that this is a mistake. "He" is always a singular term. At the same time, I grant that no expression in (L) refers to any man. These views seem very much self-contradictory: I promise I will soon enough remove the appearance of self-contradiction. That is the purpose of this chapter.

Why discuss anaphora? What bearing does this have on definite descriptions? Sentences like (L) seem to show that "he", and other sometime indexicals, are *not* singular terms. Analogues of (L) show that *definite descriptions* are not always singular terms:

(RM) "If a rich man and a poor man go to a party, the rich man will get more phone numbers."

In (RM), "the rich man" doesn't refer to anybody. It doesn't refer to Frank or Bob or Winston. So (RM), and other sentences of its ilk, seem to show that definite descriptions are not singular terms.

I very much want to show that definite descriptions *are* always singular terms. This will involve dealing with sentences like (RM). This, in turn, will necessitate our dealing with sentences like (L). Right now, we will deal with pronouns. In the next chapter, we will make the corresponding points about definite descriptions.

The apparent ambiguity of pronominal expressions

Consider the following sentences.

(1) "He [pointing to some particular individual] is a very smart man."

(2) "If John really has invented a time-machine, then he [John] is a very smart man."

(3) "If a man speaks twelve languages, he is a smart man indeed."

(4). "John has a son. He [the son] is very clever."

As Evans observes, "he" is functioning in different ways in each of these sentences. In (1), it is functioning as a demonstrative. In (2), it is functioning as what Geach calls a "pronoun of laziness": a means of avoiding repetition of the word "John".

In both (1) and (2), "he" is functioning as a singular term; it refers to some particular individual.

In (3), "he" is *not* functioning as a singular term; it seems to be functioning as a bound-variable: "for any man x, if x does what x's wife tells him to do, then x is a wise-man indeed."[147]

In (4), "he" doesn't refer to any particular person. But, surprisingly, it doesn't seem to be functioning as a bound-variable either. For (4) does not mean:

(4) There is some x such that x is uniquely a son that John has and x is clever.*

Evans was the first to recognize the significance of sentences like (4): sentences where there is a pronoun that seems to be functioning anaphorically – not demonstratively, not as a pronoun of laziness – but also seems not to be functioning as a bound-variable either. Evans refers to such pronouns as E-pronouns.

Stephen Neale pointed out that the "he" in (4) *can* be regarded as a bound variable:

(4_N) "For some x, x is a son of John's, and for any y that is identical with x, y is clever."

Thus the "he" in (4) turns out to be universal quantifier. I grant that (4_N) may be logically equivalent with (4). But, I will argue, (4_N) is semantically very different from (4); and the "he" in (4) is not a universal quantifier.

When we consider (1)-(4), the word "he" thus seems to have three or possibly *four* different meanings: indexical, surrogate of a name, bound variable, universal quantifier. But this seems implausible. Intuitively, it is very hard to believe that "he" means entirely different things in each of (1)-(4). It is very hard to believe that "he" is sometimes a singular term, sometimes a bound variable, sometimes a phonetic variant of a name, and

sometimes a universal quantifier.

We can use Kripke's (1977) classic argument to show that there is *no* semantic ambiguity here. If "he" were *semantically* ambiguous between these four meanings, then we'd expect there to be languages that had *different* words for these various disambiguations. But in *every* natural language L, the L-translation of "he" serves all four functions.[148] In Spanish or Chinese or Arabic, *one* word does duty for "he" (bound-variable), "he" (demonstrative), "he" (pronoun of laziness), and so on. Every language will assign the *same* word to each occurrence of "he" in (1)-(4).

A correct semantics must not make "he" be ambiguous. At the same time, a correct semantics must accommodate the fact that "he" *functions* in different ways - -that what it is doing in (1) is different from what it is doing in (2) or (3).

These tasks seem irreconcilable. But they are reconcilable. What changes from case to case, I will argue, is not what "he" means; what changes is the understood *domain of discourse.*

Let us say what this term means. Consider the sentence-type:

(e1) "everybody came to the party".

Depending on the context, this could mean any number of different things:

(e1) everybody in the department came to the party.
(e3) everyone in our family came to the party.
(e3) everyone who was invited came to the party.

The word "everybody" is a quantifier. Every occurrence of quantifier ranges over a certain domain: a so-called *domain of discourse.* Depending on the context, a given quantifier will be associated with different domains of discourse.

It is an open question whether this is a function of pragmatics or semantics. One *could* say that (e) always means the same thing, namely: *everybody in existence came to the party.* Nobody will deny that (e) *communicates* different things on different occasions. But perhaps this could be explained in terms of pragmatics.

For our purposes, it is enough to know that, where some expressions

are concerned, the relevant domain of discourse is subject to context-based restrictions. We may leave it open whether this context-sensitivity is semantic or pragmatic.

Some facts about indexicals

In order to come to a good understanding of pronouns, we must apply some of our points about contextual salience to indexicals.

What an indexical refers to is clearly subject to context-based restrictions on the domain of discourse. I wish to be understood aright. Obviously what "now" refers to depends on when it is uttered: the point I am making is not a mere re-statement of that platitude. What I am saying is that, depending on what the discursive background is, a token of "now" that occurs *at a given time* could pick out very different expanses of time. The word "now" – tokened at, say, 3:30 p.m. July 5th, 2005 -- could refer to this second, this minute, this day, this century, or even this geologic era. The word "here" – tokened in Washington D.C. -- could refer to this exact spot, this neighborhood, this country, or even this galaxy.

Some indexicals involve accompanying acts of ostension; others do not. If you say "that is a beautiful house", your use of the word "that" may or may not involve a demonstration: an act of pointing. If there are several salient horses, you would probably have to point to a specific one to make yourself understood. But if you said "today is a fine today", no act of pointing could possibly be involved. Kaplan refers to indexicals like "today" as "pure", and to indexicals like "that" as "impure".

It should be noted that, even with impure indexicals, an act of ostension isn't always involved. If there is only one contextually salient horse, you can say "that is a lovely horse" without pointing to anything: the word "horse" directs the auditor to look for a referent for "that" *within* a hyper-restricted domain. He will look, not just for contextually salient objects (of which there are many), but for contextually salient horses (of which there are relatively few).

But if there are many contextually salient horses, you must restrict the domain even further. One way is to *point* to some horse. If you simply say "that [no act of pointing] is a lovely horse", the auditor's search-space is too large; it includes too many horses. If you say, "that [pointing to Napoleon or Secretariat] is a lovely horse", the auditor's search space has been hyper-

restricted; the domain of discourse – the sphere within which the auditor is to look for a referent – has been made so small that there is much less room for error.

What this shows is that pointing is a way of *narrowing* the context. (It is not the only way, as we'll see in a moment.) When you point, what you are doing is to restrict the understood range of discourse: the auditor is being told to look for a referent *within* that hyper-constricted domain. So if there are several contextually salient horses, and you say, pointing to Napoleon, "that is a fine horse", the auditor knows that the "that" must refer to something within the tiny sphere delimited by your ostension.

Complex demonstratives and sortals

Nonetheless, within such a sphere, no matter how small it is, there may be different possible referents for the indexical. If I say, pointing to the space occupied by some horse, "that is a beautiful sight", the word "that" could refer to many different things: it could refer to the horse's freshly washed fur, to the manner in which the horse is leaping over some hurdle, to its musculature. It will usually be perfectly clear which of these things is being referred to. But the point is that, by itself, the act of ostension can never *completely* pinpoint an object: all it can do is to restrict the domain of discourse enough to enable the auditor to figure out what you are referring to. Oftentimes, an act of ostension is not enough to effect the needed delimitation of context. We need to add a sortal term to the indexical: we need to use a *complex* demonstrative.

Suppose you and I are looking at a field where there are many different animals. If I just point and say, "that is a lovely beast", you will have no idea whether I am referring to the sheep, the goat, or the horse. I must say "that *horse* is a lovely beast".

It must be understood that the word "horse" here is merely a way of delimiting the domain of discourse. We know, from Kaplan and others, that a demonstrative use of "that horse" is directly referential. If Alpha is the horse in question, then a token of:

(H1) "That horse is lovely"

means

(AL) Alpha is lovely.

The proposition encoded in (H) doesn't comprise the concept *horse.* The word "horse" serves to restrict the relevant domain of discourse to things that are *horses* that are contextually salient. The sortal terms in complex demonstratives are simply ways of restricting the domain of discourse.[149]

What we said about "that" applies to "he", "she", "it", and other pronouns. If you point to somebody and say "he", your act of pointing delimits the domain of discourse. The word "he" comprises a phonetically telescoped sortal: *male.* With the possible exception of "this" and "that", all indexicals contain (phonetically distorted) sortals. "today" obviously contains the sortal "day"; "now" contains the sortal "time" ("now" is short for "this time"); even "it" contains a sortal (a very inclusive one): "entity" or "thing". Even "this" and "that" probably contain sortals. Obviously "this" and "that" have different meanings; the only way to explain this is to say that the characters of "this" comprise different sortals. (Perhaps these sortals are *proximal object* and *distal object.*)

Here are the main points. First, the domain of discourse associated with a noun-phrase is subject to context-based restrictions. Second, the sortal information in demonstratives is there to restrict the domain of discourse.

In all cases, indexicals refer to the *contextually* salient object having the relevant sortal properties. This is true even of indexicals that are accompanied by demonstrations. The demonstration merely restricts the context: so the indexical *does* refer to the contextually salient instance of the sortal in question, with the qualification that the context in question is a narrowing of the context that previously obtained.

Pronouns of laziness

Now we can start to deal with the *apparent* ambiguity of "he" and other pronouns. Consider

(2) "If John really has invented a time-machine, then he [John] is a very smart man."

The word "he" refers to the contextually salient instance of the relevant sortal. The sortal in question is *male* (or *x is a male*). The contextually salient instance of it is John.

There is a wrinkle. In (2), the word "he" *does* refer to a contextually salient male. But notice that the context in question is a purely *discursive* context: the context in question does not consist of a region of space-time surrounding the speaker and auditor.

This should not alarm us. The word "context" sometimes (though much more rarely than is usually thought) refers to a region of physical space. But sometimes it refers to a region of discursive space. This does not make the word "context' ambiguous. A "context" is simply the class of entities that are, explicitly or implicitly, being spoken about. In some cases, this class coincides with the boundaries of some area of space; in other cases, it does not. Where (3) is concerned, it does not.

Consider

(1) "He [pointing to Smith] is a very smart man."

The contextually salient instance of the relevant sortal is Smith. So that token of "he" refers to Smith.

Now consider

(4). "John will have exactly one son. He [the son] will be very clever."

Here there is a wrinkle. The word "he" refers to John's son, not John. But both the son *and* John are contextually salient. The context in question is a discursive context; but, within that context, both John and his son are equally salient. This might seem to refute our idea that "he", and indexicals generally, refer to *the* contextually salient instance of the relevant sortal. For sometimes, it will be said, there are many such instances.

This is quite wrong-headed. First of all the word "he" in

(5) "John has exactly one son. He is very clever."

doesn’t *have* to refer to John’s son. It could refer to any male. You could easily imagine a context where somebody utters the first sentence, but where the "he" in the second refers to some *other* male. You and I are scientists who have invented a protein bar that dramatically increases intelligence. I say to you "John will have exactly one son". You and I then see somebody, a test subject, eating one of our protein bars. I say "he will be very clever." Here "he" refers to the test subject, not to John’s son. What has happened, obviously, is that the test-subject suddenly became the contextually salient male.

If I utter (5) to you in a vacuum, then "he" will refer either to John or John’s son. Under that circumstance, only John and John’s son will even come to close to qualifying as contextually salient. If that token of (5) is to be unambiguous, it must be made clear *which* of those two entities "he" is referring to. Only context can do this. If we have been talking about John’s recent consumption of our protein bar, then the "he" in (5) will refer to John. If we have been talking about John’s genotype, and the likelihood that any *offspring* John should have are geniuses, then "he" will refer to the son. But notice that, unless a token of (5) is to be quite indeterminate in meaning, *context* must make it clear who "he" refers to. This is another way of saying: "he" refers to the male who is salient in the relevant context. We’ve stipulated that, in (4), "he" refers to John’s son, not John. This means that, in the relevant context, John’s son is the contextually salient instance of the relevant sortal. So (4) is consistent with our analysis.

(3) is the difficult case. Here "he" doesn’t refer to *anybody.* It doesn’t refer to John or Fred or Bob. Therefore, it doesn’t refer to any contextually salient male.

But this isn’t so hard to deal with. Consider the sentence:

(6) "Smith is a man who speaks twelve languages, and he is a smart man indeed".

Just as our analysis requires, the word "he" refers to the contextually salient male, namely Smith. There is nothing mysterious about this case. Here it will help to recall what we said earlier about sentences. (6) is a phonetic variant :

(6_{PH}) "A <that Smith is a man who speaks twelve languages and he is a smart man indeed>".

The root of (6_{PH}) is the expression:

(7) "That Smith is a man who speaks twelve languages and he is a smart man indeed".

There is no difficulty explaining why the "he" in (7) refers to Smith. Smith is the contextually salient male; so the "he" refers to him. Now consider the sentence:

(8) "If Smith is a man who speaks twelve languages, then he is a smart man indeed".

(8) can be seen as built up from (7) in two phases. The first phase involves prefixing (7) with an expression that assigns truth to a conjunctive expression exactly if proposition denoted by the second conjunct is a consequence of that denoted by the first. Let "if*" be the expression denoting that function. Plainly "if*" does not differ from "if" in any *semantically* significant respect; the only differences are orthographic. So the first phase of construction yields:

(9) "If < that Smith is a man who speaks twelve languages and he is a smart man indeed>".*

In effect, we've already explained why the "he" in (10) refers to Smith: Smith is the contextually salient male.

The next phase of construction is to prefix (9) with an assertoric force-operator:

(10) A (If < that Smith is a man who speaks twelve languages and he is a smart man indeed>)".*

In effect, we've already explained why the "he" in (10) refers to Smith. So our analysis has no trouble accommodating the fact that the "he" in

(10) refers to Smith: it is a simple case of an indexical referring to the contextually salient instance of the relevant sortal.

So we've explained why the "he" in (8) co-refers with "Smith". Of course, what we said about "Smith" would hold for *any* name we replace it with. It isn't as though our analysis holds for "Smith" but not for "Dembrowski". So for any sentence of the form:

(11) "If x is a man who speaks twelve languages, then he is a clever man indeed",

we've explained why the "he" co-refers with the name occupying the place that, in (11), is occupied by the variable.

The semantics of pronouns falling within the scope of quantifiers

Before we can complete our analysis of anaphora, we have to deal with some thorny issues.

Consider the sentence

(11) "for any x, if x is a man who speaks twelve languages, then he is a clever man indeed".

Here we are tempted to say that the 'he' co-refers with the 'x' and, in general, that anaphora is co-reference induced by variable-binding.

But there is an obvious problem: the "x" doesn't refer to anything. Therefore the "he" doesn't *co*-refer with it. Where there is no reference, there is no *co*-reference.

This point must be generalized. We are often told:

Anaphora is a kind of co-reference. Suppose you say "if a man x smokes too much, he will damage his lungs". Here the word "he" co-refers with "x". But this kind of reference is different from the ordinary kind. Ordinary reference involves a relation between a sign and some object in the world. Where anaphora is concerned, the reference is entirely internal to some statement or discourse.

The problem is that a term either refers or it doesn't. If we wish, we can say that the "he" in

(12) "if a man x smokes too much, he will damage his lungs"

"co-refers" with the "x". That is, perhaps, an acceptable initial characterization: a good label for a phenomenon yet to be delineated. But strictly speaking it is false. If the "he" doesn't refer to something *actual,* then it doesn't refer at all. If we say that the "he" refers to some "discursive" entity – some entity that resides within a piece of discourse but not elsewhere – then either we're mistaking a label for an analysis or we're doing some dubious ontologizing to cover up a deficient semantic analysis.

There is no "internal" reference. Expressions either refer or they don't. If they refer at all, they refer to real things, not to "discourse-internal" things.

This brings us to a related point. Consider:

(A) "If a woman abuses her children, she forfeits any right to have her rights respected.

(B) For any woman x, if x abuses x's children, then x forfeits any right to having her own rights respected.

(A) is equivalent to (B). Because of this, we are told that, in (A), "she" is really functioning as a "bound variable". But, of course, "she" can also function as a singular term (as a demonstrative term – "she [pointing to Mary] is upset" – and as a surrogate for a name – "Mary won't leave until she's had a drink".) We are thus told that "she" is semantically ambiguous: sometimes it is a bound-variable; sometime it is a device of reference.

But the concept of a "bound-variable" is every bit as opaque as the concept of an anaphoric pronoun; by identifying anaphoric pronouns with bound-variables we are explaining the obscure in terms of the obscure;

indeed, we are explaining an obscure phenomenon in terms of itself. The concept of a "bound-variable" has generally been seen as a perspicuous and well-behaved one: presumably because that concept is used so liberally in the (from some viewpoint) most rigorous branches of learning (logic, number theory). But it is not a perspicuous notion.

In (B), the last three occurrences of "x" are bound by the same operator; we are told that, as a result, they "co-refer". But none of those occurrences refers to anything at all. None of them refers to Bob or Sally or Mary or Fred. Since they don't refer, they don't *co*-refer. As we said a moment ago, it is meaningless to speak of co-reference where there is no reference.

But, supposedly, the reason the same variable is used in those three places is, specifically, to indicate co-reference. This is the conventional wisdom. So the conventional wisdom requires that there be co-reference where there is no reference. Not an easy requirement to satisfy!

We are stuck in a vicious and dark circle. By "explaining" anaphora in terms variable-binding, we are making no progress at all.

Similar comments apply to (A). We are told that, in (A), "she" and "her" co-refer. But they don't refer at all; so they don't co-refer. We are then told that, in (A), "she" and "her" are variables that are bound by the same operator, and *that* is why they co-refer. But variables *don't* refer; it doesn't matter whether they are free or bound. By explaining anaphora in terms of variable binding, we've simply redescribed the very thing for which we seek an explanation.

How our analysis deals with these facts

Given what we just saw, it very much seems pronouns are not, at least not always, singular terms. But if we say this, then we must say – what is counter-intuitive – that "she" means one thing in "she [pointing to Mary] is angry" and an entirely different thing in sentences like (2).

I believe that all of this can be sorted out. We need only keep in mind some of the points earlier made regarding semantic rules.

Quantified sentences express propositions about classes of propositions (more accurately, about classes of open-propositions: concepts or properties, e.g. *x smokes*). Consider:

(*) "for any x, if x is a man, then that man smokes".

This says that every proposition of a certain form is true. That form is:

(**) *if x is a man, then_ that man smokes.*

The quantifier "for any x" can thus be seen as denoting a function that assigns truth to a propositional-*form* exactly if every *instance* of that form is true.

So (*) attributes truth to each instance of a class of sentences (or, more exactly, to each proposition borne by any given instance of a certain class of sentences).

Let us focus on the second occurrence of "x" in (*). That word doesn't refer to anything. Its substituends refer. When the "x" is replaced with a proper name, the result is an instance of (**). In that instance, the "that man" co-refers with the proper name, for reasons we've already discussed.

So, for reasons we've made clear, in any *instance* of (**), the occurrence of "that man" co-refers with the proper name. This point is the first step in our analysis.

Here is the next one. Quantifiers must be seen as meta-linguistic operators. "For any x" takes *expressions* for its arguments. It assigns truth to an *expression* exactly if each instance of that expression is true. Speaking approximately, (*) is equivalent to:

*(***) *if x is a man, then that man smokes* is true for all substitutions of "x".*

In (***), "that man" does not occur: what occurs is '"that man"'; in other words, what occurs is a *quotation* of "that man".

Whenever a quantifier occurs, think of what follows *not* as an expression, but a *quotation* of an expression. (The quotations are orthographically suppressed.) So in

(1) "for some x, x smokes"

think of the material following the quantifier as being in quasi-quotation marks. The second occurrence of "x" doesn't refer to anything; and neither does the first occurrence. The traditional analysis of (1) says that both occurrences *co*-refer. But how they can co-refer, if neither refers to anything?

I submit that, to a first approximation, what is really going on is this. (1) is a phonetic distortion of:

(2) "In some cases, when the 'x' in 'x smokes' is replaced with a singular term, the result is a true sentence."

In (2), "x" doesn't occur. What occurs in (2) is a *quotation* of that expression. So what occurs in (2) '"x"', not "x". Obviously the second occurrence of "x" doesn't refer.

If, as I will argue, (1) is really a distorted version of (2), then "x" doesn't occur in (1) at all. What occurs is '"x"'.

Care must be taken to avoid a confusion. Consider the sentence:

(JS) "John smokes."

What occurs in (JS) is not John – John is a person, not an expression. What occurs in (JS) is "John". Now consider the sentence:

(JSH) "John said 'hi Fred' when he entered the room."

In (JSH), for reasons just seen, what occurs is not John, but "John". It follows, obviously, that what occurs in (JSH) is not "Fred", but '"Fred"'. For exactly analogous reasons, it follows that what occurs in (2) is not "x" but '"x"'. I will now argue that (with a major qualification) (1) is really (2) in phonetic disguise. If this is right, then what occurs in (1) is not "x", but '"x"'.

Let us reconsider our paradigms:

(11) "for any x, if x is a man who speaks twelve languages, then he is a clever man indeed",

and

(12) "if a man x smokes too much, he will damage his lungs".

(11) is built out of the sentence-*form*

(11_F) "if x is a man who speaks twelve languages, then he is a clever man indeed"

We've already explained why, in each *instance* of that form, the "he" co-refers with the antecedent proper noun. (An instance of that form would be a sentence like: "if Larry is a man who speaks twelve languages, then he is a clever man indeed".) In any such instance, the antecedent proper name renders some individual contextually salient; so, given that pronouns refer to the contextually salient instances of the relevant sortals, it follows that the subsequent pronoun will refer to the previously named entity.

At the same time, the (apparent) occurrence of "he" in (12) doesn't refer to anybody; it doesn't refer to Mike or Fred or Dave.

As we discussed, this last fact is take to show that "he" is sometimes *not* a referring term – that it sometimes has a different meaning from the one it has in "if John eats too much, he'll put on weight". But that is very implausible.

Sometimes that fact is taken to show that the "he" in (11), and other quantified sentences, refers to a "discursive" entity. But that is absurd. The "he" obviously doesn't refer to Bob or Fred or any other actual male. Therefore it doesn't refer at all.

The truth is: *in (11) there is no occurrence of "he". There is an occurrence of a quotation of that expression.* (11) is a phonetic variant of:

*(11_{PH}) For all substitutions of "x", the sentence *if x is a man who speaks twelve languages, then he is a clever man indeed* is true.*

In (11_{PH}), there is no occurrence of "he"; there is an occurrence of a *quotation* of that expression.

So we are not stuck with the problem of saying what the occurrence of "he" in (11) refers to. That problem would indeed be unsolvable. But there is no problem: for there is no occurrence of "he", only of a *quotation* of that

expression. And there is no problem saying what that *quotation* refers to; it refers to an expression (more exactly, to an *occurrence* of an expression), not to some shadowy, discourse-internal entity.

There is a rather serious problem we must deal with:

You say that (11) and (12) make statements about classes of expressions. On your analysis, (11) says that every English sentence having the form

(11_F) "if x is a man who speaks twelve languages, then he is a clever man indeed"

is true. But that isn't what (12) says. What (12) says could be true even if the English language had never existed. And the Spanish translation of (11) doesn't say anything about English expressions. Consider the Spanish translation of (11).

(12_S) Cada hombre que habla doce idiomas es muy inteligente. [JK1]

That sentence would be true even if English had never come into existence; therefore, (12S) By exactly similar reasoning, for any language L, the L-translation of (12) doesn't say anything about L sentences.[150] *Therefore, (12) doesn't say anything about English expressions. So your analysis is wrong. This is a version of Church's famous "translation" argument.*

That argument mutatis mutandis shows that your analysis of all quantified sentences is wrong.

This point is an important one, and it demands that we state our analysis a little more precisely. First of all, sentence-forms are just predicates. The predicate "smokes" is really identical with "x smokes". Some predicates have multiple argument places. But this only means that they are satisfied, not by individuals, but by sequences of individuals.[151] Thus "x is jealous of y's ability to do z" is a predicate whose arguments are ordered triples, as opposed to individuals (or singletons). So predicates are sentence-forms and sentence-forms are predicates. Of course, predicates express or denote properties. "x is bald" express the property of baldness.

Given this, "for any x" is defined thus.

*Let S be any English predicate, and let P be the property that S denotes. *for any x" assigns truth to S exactly if every object has P.*

Notice that, on this definition, "for any x" denotes a function whose arguments are *expressions.* But expressions of the form "For any x…x…" do not say anything about *expressions*; they say something about the things *meant* by expressions; they say something about the properties denoted by expressions.

So, on our analysis, the thing which is *meant by*

(13) "For any x, x is an extended object"

could be true even if there were no English expressions – even if there were no expressions of any language. At the same time, the entity which "for any x" takes for its arguments is always an *expression.*

On our analysis, (13) comes to this.

The property denoted by "x is an extended object" is: x is an extended object (i.e. is that of being an extended object). Let S be any predicate of English, and let P be the property thereby expressed. "For any x" assigns truth to S exactly if everything has P. The property denoted by "x is an extended object" is: x is an extended object. So "For any x" assigns truth to "x is an extended object" is true exactly if everything has the property just mentioned, i.e. everything is an extended object.

Thus, "for any x, x is an extended object" is true exactly if everything is an extended object.

So, on our analysis, "for any x, x is an extended object" makes a statement that could be true even if the English language never existed. At the

same time, in that sentence, "for any x" takes an expression for its argument.

Thus we side-step Church's translation argument. The *Spanish* translation of "for any x" is "para cada objeto x". The semantic rule for that expression is *not*:

> *Let S be any English predicate, and let P be the property that S denotes. *para cada objeto x, S* assigns truth to S exactly if every object has P.*

Rather, taht semantic rule is:

> *Let S be an arbitrary Spanish predicate S, and let P be the property that S denotes. *para cada objeto x, S* assigns truth to S exactly if every object has P.*

So we have exactly the result we want. We want to accommodate our intuition that "he" isn't ambiguous. We *don't* want to say that "he" has one meaning in:

(1) "He [pointing to some particular individual] is a very smart man"

and a totally different meaning in:

(14) "for any man x, if x speaks twelve languages, he is a smart man indeed."

We also *don't* want to say that the "he" in (3) refers to some "discourse-internal" entity. To say that would be to create jargon to cover a lack of insight. The truth is that "he" does not occur in (3); what occurs in (3) is a *quotation* of "he". (14) comprises a *quotation* of the expression:

(15) "If x speaks twelve languages, he is a smart man indeed."

Since "he" does not occur in (14) – rather, a *quotation* of it occurs – we are spared the embarrassing task of having to say why that word refers to something in (1), but fails to refer to something in (14), even though, by all appearances, it very much seems *not* to undergo a change in meaning. We are spared that embarrassing task because it *doesn't* occur in (14); a quotation does.

We've explained why, in any *instance* of (15), the "he" co-refers with the antecedent noun. There is no need to explain why the "he" in (14) co-refers with anything; for there is no "he" in (14); rather, there is a quotation of that expression.

Chapter 27 Definite descriptions revisited

Remember what we said about indexicals. All indexicals contain sortal information; and a token of an indexical always picks out the contextually salient instance of that sortal. "He" contains a term referring to the sortal *male*; and tokens of "he" pick out the contextually salient male.

Exactly this point holds of definite descriptions. Whenever you say "the strong man" or "the mean girl in my class", the indexical picks out the contextually salient instance of the sortal (*strong man*, *mean girl in my* class).

In an utterance of:

(SN) "The senator isn’t going to be happy",

the occurrence of "the senator" picks out the contextually salient instance of the relevant sortal. If you utter (SN), and yet there is no contextually salient senator, your utterance will obviously be abortive.

Some definite descriptions always refer to the same thing, e.g. "the sum of two and three". What these expressions refer to seems *not* to be a function of contexts: they refer to the same things in all contexts. Thus, it might appear, my analysis is wrong.

I would argue otherwise. We must not confuse matters of semantics with matters of logic or metaphysics. For reasons of logic or metaphysics, only one entity can possibly instantiate the sortal concept encoded in: "the sum of two and three". Since there is *only one* conceivable instance of that sortal concept, that entity will inevitably be the *salient* instance of that concept. If you say "the sum of two and three", there is only one entity that, under any circumstances, could conceivably be the contextually salient instance of the relevant sortal. So, by default, that entity is *always* the contextually salient instance of it. So "the sum of two and three" is, semantically, just like "the strong man in the corner". Of course, in the one case, but not the other, different things can instantiate the relevant sortal information. But that has to do with the logic of the relevant sortal concepts; it doesn’t have anything to do with semantics. If we lived in a world where the sum of two and three was four in some places and was six in others, then "the sum of two and three" *would* be as context-sensitive as "the man over there".

There can be no doubt that, if we leave aside our semantic theories, and consider our intuitions, tokens of "the phi" seem to refer to the contextually salient phi. If you say "the dog looks ill", the definite description certainly *seems* to refer to the contextually salient dog; if you say "the house looks wonderful", the definite description certainly *seems* to refer to the contextually salient house. There seems to be no significant difference between expressions that are *uncontroversially* indexicals – e.g. "he", "she", "today" – and definite descriptions.

The problem is that, in some case, the definite descriptions seems not to refer

to anything. To give two examples:

(DA) "The father of each girl at the party loved her very much."
(SM) "If a strong man and a weak man arm wrestle, the strong man will win."

We encountered an exact analogue of this problem in connection with indexicals: we showed how that so-called fact did nothing to invalidate our thesis about indexicals, namely*: tokens of indexicals refer to the contextually salient instance of the relevant sortal.* We will find how the corresponding fact about definite descriptions does nothing to invalidate our thesis about them, namely: tokens of definite descriptions refer to the contextually salient instance of the relevant sortal.

(DA) and (SM) are logically equivalent to:

(DA_R) For any girl x who was at the party, if y is the father of x, then y loves x very much.
(SM_R) For any strong man x, and any weak man y, if x and y arm-wrestle, x will win.

But, intuitively, the definite descriptions in (DA) and (SM) *seem* to have the same meanings that they have when used referentially. Suppose you and a friend are at the beach, looking at a group of people; one is obviously a body-builder; the rest obviously spend no time working on their physiques. You say

(SB) "the strong man looks like he's inebriated".

The definite description is obviously functioning referentially. It would be counter-intuitive, even absurd, to suppose that "the strong man" had one semantics in (DA) and a different semantics in (SM). Definite descriptions can function in an anaphoric, quantificational, or referential manner. But *one* of these three uses is fundamental; the other two must be

understood in terms of the third. My inclination is to see the referential use as fundamental. It is very hard to believe that the *fundamental* use of definite descriptions is anaphoric or quantificational. So how are we to deal with (SM) and (DA)?

Remember what we said about quantifiers; they are metalinguistic operators; what follows them is, not an expression, but a *quotation* of an expression. "Each girl" is a quantifier. So (DA) should be rewritten thus:

*(DA_Q) For each girl x who was present, the sentence *the father of x loved her very much* is true.*

In (DA), therefore, there is no definite description, strictly speaking. Rather, there is a *quotation* of a definite description; what occurs is not "the father", but a *quotation* of that expression.

(DA), and (DA_Q) are built out of the predicate:

*(Q) *The father of x loved her very much**

For any x, *the father of x* refers to the contextually salient instance of the sortal involved. Consider "the father of the bride". The sortal in question is: *father of the bride.* Suppose you say:

(FB) "the father of the bride looks exhausted"

If Smith is the contextually salient instance of the sortal, then the definite description refers to Smith. The sortal associated with "her" is: *female.* So in any instance of (Q), "her" refers to the contextually salient female. In an utterance of:

(FJ) "Jenny's father loves her very much"

"her" refers to Jenny. And in a context in which there occurs an utterance of:

(LV) "The father of the bride loved her very much".

the bride is made contextually salient by the occurrence therein of "the bride." The bride is the contextually salient female. Thus "her" co-refers with "the bride". Given any instance of (Q), our analysis makes it easy to explain why "her" co-refers with "the bride".

What occurs in (DA_Q) is not (Q), but a *quotation* of (Q). (DA_Q) makes a statement about (Q) itself; (Q) is mentioned, not used; so what occurs in (DA_Q) is a *quotation* of (Q). Thus, what occurs in (DA_Q) is not an occurrence of "the father", but a *quotation* of that expression. Therefore there is no need to say what the occurrence of "the father" in (DA_Q) refers to: there is no such occurrence; what occurs is not "the father", but a quotation of that expression.

Here we must be very careful. Even though (DA_Q) mentions the expression "the father of x loved her very much", the proposition encoded in (DA_Q) is not about an English expression. It is true that the material following the quantifier is a quotation. But the proposition encoded in (DA_Q) is not metalinguistic. We must remember how quantifiers are defined. "For each x" is defined thus.

> *Let S be the quotation of any English predicate and let P be the meaning of the disquotation of S. *For each x, S* is true exactly if everything instantiates P.*

So even though "for each x" takes *quotations* of expressions for its arguments, nonetheless sentences of the form "…for each x…" are not about expressions; they are about the things meant by expressions. That is why the proposition encoded in sentences of the form "…for each x…" typically have nothing to do with expressions. That is why the Spanish or Albanian

translation of such a sentence doesn’t say anything about English expressions.

(DA_Q) is just a phonetic variant of (DA). So what occurs in (DA) is not "the father", but a quotation of that expression. Thus, we are spared the embarrassing task of saying what "the father" in (DA) refers to. We are spared having to say that it refers to some "discourse-internal" parent; and we spared having to say "the father" is not a referring term. In (DA), there is no occurrence of that expression: a *quotation* of it occurs, not that expression itself. (DA) makes a statement about the class of sentences having the form:

*(Q) *The father of x loved her very much**

The expression "the father" occurs in each sentence falling into that class. So in any such sentence, "the father" refers to somebody; and a correct semantics must say why it refers to that person. We’ve already explained this. But in (DA) itself, "the father" does not occur; we are thus spared the embarrassing task of having to say what that occurrence refers to. That occurrence doesn’t exist. There is no definite description in (DA); there is a quotation of such an expression.

It is absurd to suppose that "the strong man" and "the father of the bride" are semantically *ambiguous* between referential, anaphoric, and quantificational meanings. Kripke’s (1977) argument shows this. At the same time, it is absurd to say that, in (DA), "the father of x" refers to somebody. It doesn’t refer to any particular individual; therefore it doesn’t refer at all. Nothing containing a variable, whether that variable is bound or free, can refer.

This has tempted many, including Evans, to say that definite descriptions don’t refer. But this is radically counter-intuitive.

None of these desperate measures need be taken. "The phi" always refers to the contextually salient phi. When we *seem* to have a non-referring use of a definite description, what is *really* occurring is a quotation of a definite description, not an actual definite description.

There is a nicety that we must deal with. Consider the sentence:

(^) "John owns a donkey. He beats it often."

Here the tendency is to say that "it" is functioning *anaphorically* – as a device, not of genuine reference, but of "internal" reference. The tendency is to see the "it" in (^) as being in the same category as the "he" in "if a man smokes, he puts his health at risk". I do not think this is so. What is going on in (^) is just what Evans, Bach, and others thought. The first clause *describes* something. The second refers to the thing described. Things are sometimes made cognitively, and therefore referentially, available to us through existence claims. (^) is simply an instance of this phenomenon.

People and things are often identified descriptively. "Socrates was the philosopher who drank hemlock". "Mercury is the planet closest to the sun". We are obviously able to *refer* to things to which we are introduced through descriptions. (The descriptive information merely fixes the referent; it doesn't give the meaning. That is why our modal intuitions are not corrupted; and that is why "Socrates" and "Mercury" designate rigidly.)

I believe that sense-perceptions encode descriptive information, and that even when one acquires a concept of something through-sense perception, knowledge of a description is still involved: *the guy standing over there, the dog I'm looking at right now, the man in the red car*. Ostensive definitions involve sense-perceptions; and, for this reason, it is thought that ostensive definitions are not descriptive. But it is precisely *because* ostensive definitions involve a sense-perception of the thing ostended that they are, in effect, descriptive. When you point to something and say "*that* is atrocious", your ostension works only in so far as the auditor has a perception of the thing indicated. That perception will necessarily be descriptive. From his viewpoint, it will be given by a description like *the man over there, next to the tree*. So ostensive definition is not, I believe, to be contrasted with descriptive definition: the former is a species of the latter.

I believe that, E-pronouns, like the "he" in (^), are not cases of so-called anaphoric reference. (I say "so-called" because, as we've seen, "anaphoric reference" is an oxymoron.) They are cases of garden-variety *actual* reference. An object is indicated (it is indicated by a description, not an ostension). The pronoun simply refers to the indicated object.

The alternative is to take Neale's position. We can say that the "he" in (^) is a bound-variable. This means that (^) is synonymous with:

> *(^^) "There is some x such that x is uniquely a donkey owned by John, and for any y such that y is identical with x, John beats y."*

But if Neale’ proposal is right, then "he" is sometimes a universal quantifier, in addition to being a singular term. For reasons discussed earlier, this is not very plausible. I think it better to take the Evans-Bach position, according to which the "he" in (^) is a singular term that refers back to the thing just described.[152]

Chapter 28 Can reference be reduced to quantification?

Many semanticists think that definite descriptions are quantifiers. Everybody agrees that definite descriptions have referential *uses.* Those semanticists try to show that this fact is compatible with their being quantifiers, at the level of literal meaning. Suppose there is exactly one contextually salient man wearing a red hat. You say:

(RH) "the man wearing the red hat is a lawyer".

Of course, (RH) can certainly be *used* to communicate a proposition like:

(RHO) Smith is a lawyer.

The definite description can be *used* referentially. But according to the Russellian, (RH) always means:

(RHR) Somebody x is uniquely a contextually salient man who is wearing a red hat, and x is a lawyer.

Russellians say that, in some circumstances, it will be obvious from context that, if you uttered (RH), you were trying to communicate a claim like (RHS). So, even though any token of (RH) has (RHR) for its literal

meaning, context will make it obvious that what the speaker is trying to communicate is some claim like (RHS). Basically, when we take pragmatics into account – when we consider the interactions of literal meaning and context – the Russellian analysis is perfectly compatible with the fact that definite descriptions are *used* as devices of reference.

This is not tenable. The idea that Russell's theory has any plausibility is based on a failure to ask, and therefore to answer, the question: is Russell's theory a piece of conceptual analysis or of semantic analysis? Let us consider each possibility. First, suppose he is doing conceptual analysis. In that case, his theory amounts to this.

> *Let S be an arbitrary sentence of the form "…the phi…". Let P be the proposition semantically encoded in "…the phi…". In that case, P is logically equivalent with: something x uniquely has phi and…x…P isn't necessarily the proposition that is semantically encoded in "…the phi…" But it is logically equivalent with that proposition.*

When you give a *conceptual* analysis of a sentence, you are giving conceptually necessary and sufficient conditions for the truth of the proposition encoded in that sentence.

If the proposition semantically encoded in "the phi has psi" is something of the form: *O has psi*, then Russell's' theory is a failure, if taken as a conceptual analysis. No proposition of the form *O has psi* is logically equivalent with any proposition of the form *something x uniquely has phi and any such thing has psi.* Supposing that Russell's theory is a piece of conceptual analysis, then "the phi" cannot, at the level of literal meaning, be a referring a term. For if "the phi" is a referring term, then the proposition encoded in "the phi has psi" will bear no resemblance to *exactly one thing x has phi and x has psi.* So *semantically* "the phi" must be a certain kind of quantifier if Russell's theory is to be a tenable *conceptual* analysis. It is obvious what kind of quantifier it must be. So Russell's theory is a tenable *conceptual* analysis only if it is a tenable *semantic* analysis. We cannot exonerate Russell's theory by saying that it is not a piece of semantic analysis. A semantic analysis is precisely what it is.

A semantic analysis gives literal meaning. So Russell's theory amounts to this.

The literal meaning of

(i) "the dog looks ill"

is

(ii) something x is uniquely a contextually salient dog, and x looks ill.

(ii) is a proposition, not a sentence. But it is very easy to produce a sentence that has (ii) for its literal meaning:

(iii) "something x is uniquely a contextually salient dog, and x looks ill."

So Russell's theory is really this:

RT

(i) and (iii) have the same literal meaning; they are synonymous. In general.

"the phi has psi"

and

"something x is uniquely a contextually salient phi, and x has psi"

have the same literal meaning; they are synonymous.

It is very easy to show that (i) and (iii) are *not* synonymous. Two sentences are synonymous exactly if, semantically, they are distinguishable. So the only differences between synonymous sentences must be entirely orthographic (or acoustic). So if Russell's theory is right, then (i) and (iii) are

mere phonetic variants. They differ from each other as a Southerner's pronunciation of a sentence will differ from a Northerner's. But the actual differences between (i) and (iii) are much deeper than that. Everybody – Russellians and referentialists alike – will agree that (iii), but not (i), is appropriate as answer to the question "how many contextually salient dogs are there?". And everybody – Russellians and referentialists alike – will agree that (i), but not (iii), is appropriate as an answer to the question "how is the dog doing today?" Russellians say that pragmatics – facts about implicature and the like – are responsible for these deviations.[153] But if (i) and (iii) were really synonymous, then they could no more deviate from each other, in respect of appropriateness, than a Southerner's token of a sentence could deviate, in respect of appropriateness, from a Northerner's. Mere orthographic, or acoustical, differences are not going to generate a need to invoke pragmatics.

Suppose you wanted to *assert* how many contextually salient dogs there were. You couldn't *possibly* use (i) to do so. Russellians agree with this. No Russellian has ever maintained that (i) is even remotely comparable to (iii) as a means of making such an assertion. Russellians blame this difference on pragmatics. But if (i) and (iii) are synonyms, then they are mere phonetic variants of each other. Pronounce (iii) any way you wish; give it any phonetic realization humanly possible. No such realization will make (i) be pragmatically anything like (iii). No matter how you pronounce (i) and (iii) – so long as your words are recognizable and so long as you are uttering those sentences and not different ones – the pragmatic differences between (i) and (iii) will remain firmly in place. If (i) and (iii) were mere synonyms, and were mere phonetic variants, it would be possible to efface these differences through phonetic (acoustical) adjustments.

There is the obvious reply:

The sentence

(i) "lawyers make a lot of money"

is synonymous with

(ii) "shysters make a lot of money".

There are no *semantic* differences between them. But there are pragmatic differences; there are differences in respect of what they communicate.

The same is true of "but" and "and".

(iii) "Tim is strong but Tim does not lift weights"

is synonymous with

(iv) "Tim is strong and Tim does not lift weights".

Anything that can be inferred from (iv) can be inferred from (iii), and vice versa. But (iii) and (iv) are pragmatically very different. A token of (iii) indicates that the speaker sees no tension between the two conjuncts; a token of (iv) indicates that the speaker sees just such a tension.

Exactly similar arguments show that "f**k" and "make love", "hooker" and "courtesan", "janitor" and "custodial engineer", are synonyms even though they differ pragmatically.

§ This is a good place to weigh on this old issue. Frege did indeed say exactly that about "but" and "and", and also, at least by implication, about "shyster" and "lawyer". But Frege's point has never really struck anybody as a faithful representation of natural language. And, I would argue, it is demonstrably *not* accurate.

The difference between "but" and "and" is different from that between "shyster" and "lawyer" (and "f**k" and "make love"). We must deal with these cases separately.

Imagine a three year old who thought that "but" and "and" were interchangeable. Surely we would diagnose him with a language disorder. Imagine somebody learning English who thought the same about "shyster" and "lawyer". We would say that he hadn't quite mastered the English language. Some fact about English semantics eludes him. Obviously the same is true if somebody sees no difference among "f**k" and "have sexual intercourse with" and "make love to". That person has not quite caught onto the finer points of English semantics.

Suppose you are talking with Dave. Dave says:

"Everybody who is strong lifts weights. It simply isn't biologically possible to become strong in any other way."

In response, you say:

(iii) "Tim is strong but Tim does not lift weights"

For the sake of argument, suppose you are right: Tim is strong and he doesn't lift weights. Given the circumstances, there is a tension between both conjuncts: given what Dave said, the first conjunct is in tension with the second. Some proposition P has been asserted such that the first conjunct is, and the second conjunct is not, consistent with P. P is Dave's assertion that only weight-lifters are strong. There is no *logical* tension between (i) and (iii); they are both true. But given Dave's statement, they are obviously in tension with each other. Your use of "but" is appropriate precisely because of that tension.

Let us now consider a case where both conjuncts are true but "but" is *in*appropriate.

Dave: All primes greater than two are odd.

Bob: The number seven is a prime greater than two. But seven is odd.

Bob's response is very strange. The requisite tension isn't there.

Before we can say what, precisely, is meant by "but", we have to deal with a nuance. Is Bob's statement *false*? Or is it rather abortive? My inclination is to see it, not as false, but as involving a false presupposition, and therefore as being abortive. But this is an open question. What everyone agrees on is that Bob's statement is *somehow* deviant. If the deviance is due to falsity, then the meaning of "but" is given by one semantic rule. If the deviance is due to abortiveness, then its meaning is given by some other semantic rule. But in *either* case, I will now argue, the rule for "but" is different from that for "and".

Remember that "and" denotes a function that assigns truth to a pair

of sentences exactly if both members of that pair encode true propositions. The rule for "but" is similar, but noticeably distinct. If Bob's statement is *false*, then the rule for "but" will approximate to this:

> *(But$_F$) For any sentences S_1 and S_2, a token of *S_1 but S_2* is true exactly if both conjuncts are true and, in the context of tokening, the truth of the first is in tension with the truth of the second.*

If (But$_F$) is correct, then "but" denotes a function F that assigns truth to a pair of sentence tokens exactly if both tokens are true and, in the context of tokening, there is a tension between the truth of the first and the truth of the second.

Many, including myself, think that Bob's statement is abortive – not appropriately characterized *either* as true or false (though it obviously has true components). If that is right, then "but" is defined thus:

> *(But$_A$) Let S_1 and S_2 be any two sentences. Let s_1 and s_2 be tokens of those sentences that occur in the same context. If each of s_1 and s_2 is true, and there is a tension between the truth of the one and that of the other, then a token of *S_1 but S_2* is true. If there is no such tension, or one of the conjuncts is not true, then that token is abortive.*

If (But$_A$) gives the semantics for "but", then "but" denotes a function F that is defined as follows. For any two sentence-tokens t_1 and t_2 that occur in the same context, if t_1 and t_2 are both true, and *if*, in the context in question, there is a tension between the truth of the one and that of the other, then F assigns truth to the pair $<t_1, t_2>$. If there is no such tension, then F doesn't assign anything; it is undefined in that case.

It is patently obvious that "but" and "and" are *not* interchangeable, and that this non-interchangeability has a semantic basis. It is an open question whether deviant uses of "but", like Bob's result in falsity or in abortiveness. But in either case, as we've seen, the semantic rule for "but" is markedly different from that for "and".

A similar point holds for pairs like "shyster" and "lawyer", "hooker" and "courtesan", and the like. Actually, the differences between (say) "lawyer" and "shyster" is easier to deal with than that between "but" an "and".

Let us start with "shyster". Setting aside questions about literal meaning, what is *understood* by sentences like

(*) "Bob is a shyster",

is (i) that Bob is a lawyer and (ii) lawyering is a dishonorable profession.

For the sake of argument, suppose that, without exception, all lawyers make a lot of money. But suppose also that lawyers are, without exception, utterly honorable and good people, and that the profession of lawyering is beyond reproach. Now consider the sentence:

(L) "Shysters make a lot of money. "

Is (L) true or not? Here people have two different reactions. Some will say that (L) is false. A condition for its truth hasn't been met. Others will say that a *pre*-condition for its truth hasn't been met, and that it is therefore neither true nor false. (For the record, my own view is the latter.) Everybody agrees that, under the circumstances described, (L) is defective in some way or other; everybody agrees that it would not be quite right to say "yes" to (L). Some people say this inappropriateness corresponds to the falsity of (L); others say that it corresponds to the abortiveness of (L). In *either* case, the semantic rule for "lawyer" is going to be different from that for "shyster."

Let us start with the term "lawyer". To simplify things, let us start with the predicate "x is a lawyer". The term "lawyer" has other uses; for example, it can function as a part of a quantifier ("some lawyers are evil"). But what we will say about "x is a lawyer" is easily fitted to these other functions.

The semantic rule for "x is a lawyer" is, approximately, this:

*(law) For any x, a token of *x is a lawyer* is true exactly if x is a lawyer.*

So "x is a lawyer" denotes a function that assigns truth to an object exactly if that object is a lawyer.

Let us now talk about the term "shyster". If

(L) "Shysters make a lot of money ",

is *false*, then the semantic rule for "x is a shyster" is this:

*(shy_F) For any x, a token of *x is a shyster* is true exactly if x is a lawyer and lawyering is a dishonorable profession.*

If (L) is *abortive* – if a precondition, as opposed to a condition – for its truth hasn't been met, then the semantic rule for "x is a shyster" is at least approximately:

*(shy_A) If the profession of lawyering is typically a dishonorable one, then for any x, a token of *x is a shyster* is true exactly if x is a lawyer. If the profession of lawyer is not typically a dishonorable one, then *x is a shyster* is abortive, for any x.*

If (shy_F) is right, then for any x, "x is a shyster" denotes a function that assigns truth to x exactly if x is a lawyer and lawyering is, typically, a dishonorable a profession.

If (shy_A) is right, then *if* lawyering is typically a dishonorable a profession, then for any x, "x is a shyster" denotes a function that assigns truth to x exactly if x is a lawyer. If lawyering is not, typically, a dishonorable profession, then "x is a shyster" is undefined (abortive) for all values of x.

Conventional "Implicature" as Wide-scope Semantics

Consider the dialogue:

A: Where can I buy gas?
B: There is a gas-station around the corner.

B's words do not *literally* mean: *you can buy gas around the corner.* But that is the message *conveyed* by B's words. As Grice put it, B has

implicated that A can buy gas around the corner; B's words carry the *implicature* that A can buy gas around the corner.

The sentence "there is a gas station around the corner" doesn't *necessarily* have that implicature. Whether the implicature holds is a function of the conversational context. So Grice referred to it as "conversational implicature".[154]

Conversational implicature is cancelable. This is a corollary of the fact that conversational implicature is a function of the conversational context:

A: Where can I buy gas?

B: There is a gas station around the corner. But it has been closed for years. So you can't buy gas there.

With those extra two sentences, B has changed the conversational context, and thus nullified the implicature.

A moment ago we argued that "lawyer" and "shyster" differ *semantically.* Grice denied this: he said that semantically "lawyer" and "shyster" are identical: they are synonyms.

But Grice also granted – what cannot be denied – that "shyster" and "lawyer" are not interchangeable. He said that this non-interchangeability had to do with pragmatics -- with *implicature.*

There is an obvious difference between *this* kind of implicature – if implicature be the right term -- and the kind mentioned a moment ago. The value-judgment attaching to "shyster" is not a function of the conversational context. Any use of "shyster" bears the message that lawyering is a dishonorable profession: that message is endemic to the word. Grice recognized this. Grice refers to this kind of implicature, if implicature be the right word, as "conventional implicature".

Conventional implicature is not "cancelable". This is a corollary of the fact that it is not context-dependent.

So, in Grice's view, "hooker" and "courtesan" are synonyms; they differ, not in semantics, but in implicature. The same is true of "f**k" and "make love", "janitor" and "custodial engineer", and so forth. Astonishingly, Grice held exactly this about "but" and "and": semantically, they are

identical; but they differ in "conventional implicature".

I must take exception to Grice's analysis. So-called conventional implicature isn't implicature at all; it has nothing to do with pragmatics. The term "conventional implicature" is really just a very misleading label for an aspect of semantics: one that isn't reflected, or is reflected only subtly, in differences in truth-conditions – or in truth-*pre*conditions, rather. Suppose Timmy thinks that "shyster" is interchangeable with "lawyer". Obviously Timmy doesn't quite know what "shyster" means; there is something about the meaning, the semantics, of that expression of which he is ignorant. If we wish, we can refer to that something as a "conventional implicature". But then we are using the term "implicature" to cover a bit of *semantic* terrain. Saying that "shyster" and "lawyer" – and "f**k" and "make love", and so on – differ in "implicature" is just a way of passing the buck, of dodging the difficult task of stating *how* exactly they differ. It is, at best, just a label for a problem; at worst, it is an erroneous diagnosis. The differences between (say) "shyster" and "lawyer" are endemic to those words; they are not contextual. For the sake of argument, suppose we agree with Grice. Suppose we say that, in some cases, "implicature" denotes something context-sensitive and, in other cases, it denotes something context-insensitive. In that case, we are using a single word to cover two entirely different phenomena.

The conventional reasoning is something like this:

> *Differences in conventional implicature seem to amount to differences in innuendo – differences in what is suggested -- and not differences in what is said. What is said falls within the scope of semantics. What is suggested – innuendo and the like – is a matter of pragmatics. So the innuendo attaching to words like "shyster" is, we are told, a matter of pragmatics, not of semantics.*

I think this reasoning is misguided. I grant that "shyster" differs specifically in innuendo, and nothing else, from "lawyers". But I think that it is misguided to chalk this up to pragmatics. Facts about innuendo oftentimes reflect subtle facts about semantics. Conventional "implicatures" are obviously rooted in facts, albeit subtle ones, about the *semantics* of words. It is obviously part of the very meaning of "shyster" – and "f**k" and "hooker"

– that they differ in innuendo from "lawyer", "make love", and "courtesan".

The idea that so-called "conventional implicature" – "tone", "coloring" – is not a matter of semantics is based, in my view, on a conflation of two distinct notions: the notion of *truth-conditions*, on the one hand, and *semantics* on the other. The *truth*-conditions of:

(shy) "Shysters make a lot of money",

are *not* so different from those of

(law) "Lawyers make a lot of money".

But remember the debate between Strawson and Russell. Both agree that

(IB) "The inventor of bifocals snored"

is true exactly if there is (or was) exactly one inventor of bifocals x and x snored. But the difference between their views lies in where, in the relevant semantic rules, they put the "means that" operator. Strawson gives it narrow scope; the result is that what (IB) *literally says* is *de re* about a certain individual: the rest is conveyed by innuendo, by implicature. As we saw, a consequence of Strawson's theory is that (IB) *communicates* information about the concept *inventor of bifocals.* But it is also a consequence of Strawson's theory that it is communicated *non-semantically*; it is communicated by innuendo. The *fact* that (IB) communicates information about that concept is obviously to be traced to facts about the *semantics* of the expression. Remember that the innuendo in question is a by-product of the fact that the *semantic* rule for (IB) is:

If somebody x is uniquely an inventor of bifocals then "the inventor of bifocals snored" means: x snored. If there is no such person, then that sentence doesn't mean anything.

So while the literal meaning of (IB) is simply *Franklin snored*, nonetheless information about the concept *inventor of bifocals* is conveyed by innuendo: *but this fact has a semantic basis; the explanation is to be found*

in facts about the semantics of the definite description. We cannot -- we must not -- always dump innuendo on pragmatics. In some cases, of course, innuendo is always a matter of pragmatics: but not always. Notice that the innuendo attaching to "the inventor of bifocals" is inseparable from that expression; it is not cancelable. But that innuendo is, without any doubt, to be traced to facts about the semantics of the expression in question. In some cases innuendo falls within the province of pragmatics (conversational implicature). But if the innuendo is inseparable from the expression, we should think twice before giving our dirty semantic laundry to pragmatics.

Any semantic rule contains an operator like "means that" or "refers to". Information that is *presupposed* by correct use of the expression in question is given *wide-scope* with respect to that operator. Presupposed information is not asserted, but it is still communicated; it is not *said* but is communicated by innuendo. So conventional implicature is *wide-scope semantic information.* A precondition for non-abortive use of (shy) is that lawyering be a dishonorable profession. *Given* that presupposition, (shy) and (law) say exactly the same thing; they have exactly the same truth-conditions. (shy) does not *assert* that lawyering is a dishonorable profession; it *presupposes* it. That is why (shy) is semantically different from:

(ldp) "Lawyering is a dishonorable profession, and lawyers make a lot of money."

(ldp) says what (shy) only presupposes. This is because, in the semantic rule for (shy), the business about lawyering being dishonorable is given wide-scope with respect to the "means that" operator, whereas in (ldp) is it given narrow-scope.

To sum up, "conventional implicature" is wide-scope *semantics.* This means that differences in conventional "implicature" are *semantic* differences. This fact utterly nullifies the above criticism of our attack on Russell’s theory.

To sum up differently, conventional "implicature" is a semantic phenomenon. But it is not a truth-conditional phenomenon. It is a truth-*pre*conditional phenomenon. When you say "shysters make a lot of money", you are *presupposing* that lawyering is a dishonorable profession. A

precondition for the non-abortiveness, and therefore the truth, of your utterance is that lawyering be dishonorable. But you are not *saying* that lawyering is dishonorable. Truth-*preconditions* corresponding to wide-scope semantic content. The points we just made about "shyster", "hooker", and the like correspond quite closely to the points we made about definite descriptions. If you say "the inventor of bifocals was tall", you are not *asserting* that there was exactly one inventor of bifocals; you are presupposing it. This is because the descriptive content associated with "the inventor of bifocals" is given wide-scope in the semantic rule for that expression. That descriptive content is indeed part of the semantics of that expression: that is why "the inventor of bifocals" *conveys* that there was exactly one inventor of bifocals. But, in the relevant semantic rule, that content is given wide-scope: that is why it sounds strange to say "that is false: there were two inventors of bifocals" in response to "The inventor of bifocals snored". The semantic rule for "shyster" includes information to the effect that lawyering is dishonorable. That is why "shysters make a lot of money" conveys that lawyering is dishonorable. But that information is given *wide* scope in that rule: it is not a condition, but a *precondition*, for the truth of "shysters make a lot of money" that lawyering be dishonorable. So in uttering that ,you are not *asserting* that it is dishonorable; you are presupposing it. That is why it sounds strange to say "you are wrong: lawyering is not dishonorable" in response to an utterance of "shysters make a lot of money".

Referential-attributive

Keith Donnellan[155] observed that definite descriptions can be used in two very different ways: "referentially" and "attributively".[156]

Any definite description encodes sortal information: "the tall man", "the most important equation", "the dumbest thing I ever did". But this information can play very different roles. In some cases, it is used only to *identify* the individual about whom one wishes to make a statement: the descriptive information does not itself figure in the statement. In other cases, the descriptive information is itself a constituent of the statement made.

Suppose you and a friend are looking at a group of women. One of them is wearing a blue wig. You say "the woman wearing the blue wig is a

partner at my law-firm". Here there is some specific woman x such that you have x in mind and you are making a statement about x. The definite description – "the woman wearing the blue wig" – encodes descriptive information. But you are using that information only as a means of picking about some particular individual; the statement you are making is about that individual, not about the concept *woman wearing a blue wig.* The definite description is being used to refer to that woman; it is being used "referentially". You are not saying anything about the concept *woman wearing a blue wig*; you are using that concept to *identify* the thing about which you are making a statement.

In some cases, the descriptive information actually figures in the proposition being made. Suppose you have no idea who the president in 2045 will be. You say

(pr) "the president in 2045 will have a very difficult job".

When you utter (pr), there is no *specific* individual x about whom you wish to make a statement. You are not saying that Harry will have a hard job, or that Frank or Fred or any other specific individual will have a hard job. You are making a statement about the very *concept* of being a president in 2045: you are saying that any instance of that concept will be an instance of some other concept (*will have a hard job*).

The referential-attributive distinction is a legitimate one. Whether it is a matter of semantics or merely of pragmatics is another matter: one we will weigh in on shortly.

Everybody thinks that, at last *prima facie,* Russell's theory is in conflict with the fact that definite descriptions can be used *referentially.* After all, Russell's theory just *is* the thesis that definite descriptions are not singular terms. But it is generally taken for granted that Russell's theory is perfectly consistent with *attributive* use of definite descriptions. It is always assumed that, if you use "…the phi…" in an attributive manner, you are using that sentence to *assert*:

Something x uniquely has phi and…x…

So it is always assumed that when somebody says:

(pr) "The president in 2045 will have a very difficult job".

They are *asserting*:

(pra) somebody x will be a unique president in 2045 and x will have a very difficult job.

This is all false. Even when definite descriptions are used *attributively*, existence is being *presupposed* no less than when they are being used referentially. Russell's theory is as *inconsistent* with the attributive use of definite descriptions as it is with their referential use.

Suppose you and Bob are discussing the future of American politics. Bob says that the U.S. is undergoing radical political changes. He seems to be suggesting, though he doesn't say it outright, that soon the U.S. will not be run by a single president, but by a committee of presidents. You ask Bob "will there be a single U.S. president in 2045 or not?" Bob says:

(i) "In 2045 there will be exactly one U.S. president; he will have a lot to deal with".

(i) seems like a fairly appropriate response to your question. Now suppose that, instead of saying (i), Bob says:

(ii) "In 2045 the U.S. president will have a lot to deal with".

Bob's statement is deviant. He seems to be presupposing the very thing that, under the circumstances, he should be asserting. This is not a nuance. If Bob thought that (ii) was *as* responsive as (i) to your question, he would be guilty of serious linguistic incompetence. Here we are dealing with linguistic competence, not with unresponsiveness to the unstated protocols governing interpersonal discourse.

A slight detour will put this in context. Suppose you ask Fred "I'm running low on gas. Is there a gas station nearby?". Fred says: "there is a gas station around the corner." You drive around the corner; there is a gas-station there. But it has obviously been closed for years. There are two possibilities here. First, Fred was intentionally being obstreperous; he knew that the gas-station was closed, and he intentionally misled you. Second, Fred is stupid; he didn't mean to mislead you; he tried to answer your question; but, because he is so stupid, he answers the question you were *literally* asking, not the one you were *really* asking. In the first case, Fred is clearly not guilty of any kind of incompetence; he is simply being immoral. In the second case, Fred is guilty of *some* kind of incompetence. But is it *linguistic* incompetence? Is the problem that Fred doesn't have a mastery of the English language? No. The problem is a more general form of *cognitive* incompetence; his linguistic competence is unexceptionable. Fred's torpid mind cannot track the unstated understandings that bridge the gap between literal meaning and speaker's meaning. But Fred is not guilty of *linguistic* incompetence, at least not in this context.

But suppose Bob says (ii), instead of (i), in response to your question. That *does* show linguistic incompetence. That is not a function of garden-variety stupidity; that shows a deficit in semantic knowledge. We are not dealing with garden-variety obtuseness.

There is no doubt that any utterance of (ii) must be taken *attributively*. (In any case, it must be taken attributively until it is *known* who the president in 2045 is. Thereafter, that sentence could be given either referential or attributive readings.) But we've just seen that (i) and (ii) are quite definitely *not* interchangeable. So (i), even when meant attributively, is by no means interchangeable with (ii).

Another example might help. Consider the following dialogues:

Me: There is a unique highest prime, and that number is greater than 100."

You: "That is false. There is no highest prime.

In this dialogue, your response is unexceptionable. It is in no way deviant. Now consider a slightly different dialogue:

Me: The highest prime is greater than 100.
You: That is false. There is no highest prime.

In the second dialogue, your response is decidedly less appropriate. In the first dialogue, the definite description is being used attributively. In that dialogue, I am not saying that 134 or 896 or any other specific number is the highest prime. Nonetheless, even though my sentence is being used attributively, it is clearly not interchangeable with "there is a unique highest prime, and that number is greater than 100. This isn't a matter of pragmatics. The following sentences are *not* interchangeable:

(*) "There is a unique highest prime, and that number is greater than 100."

(**) "The highest prime is greater than 100".

It doesn't matter if (**) is read attributively; the two are still not interchangeable. This isn't a matter of pragmatics. Anyone who thought they were interchangeable would be guilty of a serious linguistic deficit, not of garden-variety obtuseness or insensitivity to innuendo. Russell's theory is as false for attributive uses of definite descriptions as it is for referential uses of them. This obviously favors the view that definite descriptions are singular terms.[157]

The attributive function of definite descriptions

We still have a problem. When "…the phi…" is read attributively, it is *not*, we have seen synonymous with "something x uniquely has phi and… x…" Nonetheless, when "the phi" is functioning attributively, it is surely not referring to anyone. Consider:

(ii) "In 2045 the U.S. president will have a lot to deal with".

Here the definite description is not referring to Smith or Jones or Brown. The proposition communicated is about the concept *president of the U.S. in 2045*; it is not about this or that particular individual. How is a referentialist to deal with this?

Kuczynski (2004) argues that Russell's theory is false; and he tries to deal with the referential-attributive distinction within a referentialist framework. He writes:

> *"When '…the phi…' is functioning attributively, the definite description is seen as referring to an object described in an understood, antecedent existence claim."*

What Kuczynski means is this. Consider (ii). In Kuczynski's view. The proposition semantically encoded in (ii) has the form:

Alpha will have a lot to deal with.

Kuczynski also grants that what (ii) communicates is not a proposition about any particular individual; it is a proposition about the

concept *president of the United States in 2045*. He says that, when a definite description is functioning attributively, that is because it is preceded by a *tacit* existence claim, and the description refers to the person *described* in that claim. So (ii) is preceded by a tacit existence claim, something like:

(iii) Somebody x will be a unique president of the United States in 2045.

The definite description in (ii) *refers* to x: it refers to the person described in the antecedent existence claim.

Kuczynski's analysis harmonizes with a point that Kripke makes and that we generalized. Reference is always *fixed* by existence claims. Suppose you ask "who was Socrates?" and you are told "Socrates was the protagonist in most of Plato's dialogues." Here what you are being told is: somebody x was uniquely a protagonist in most of Plato's dialogues, and "Socrates" refers to x. We observed that, even in cases of *ostensive* definition, reference is fixed by existential information. (The existential information is encoded in the perceptions accompanying the ostension; it is not verbal, at least not entirely.) Reference is thus *always* parasitic on understood, antecedent existence claims; and Kuczynski's analysis seems consistent with this.

I do not know whether the solution just described is correct. But I would suggest that there is a simpler one. For the sake of argument, suppose that the referentialist is right. In that case, the semantic rule for "the president of the U.S. in 2045" is this:

If somebody O is uniquely a president of the U.S. in 2045, then

"…the president of the U.S. in 2045…"

encodes the proposition:

…O…

If there is no such person, then

"…the president of the U.S. in 2045…"

is abortive.

As we've seen, an immediate consequence is that, given anyone who knows the semantics of "the president of the U.S. in 2045", that person will know that

"…the president of the U.S. in 2045…"

is true exactly if somebody O is uniquely a president of the U.S. in 2045 and…O…

In particular, any such person will know that:

(ii) "In 2045 the U.S. president will have a lot to deal with".

is true exactly if somebody O is uniquely a president of the U.S. in 2045 and O has a lot to deal with.

At the same time, if the referentialist is right, then (ii) encodes a proposition of the form

(*) Alpha will have a lot to deal with.

So, if the referentialist is right, then (ii) is synonymous with some sentence of the form

(**) "So and so has a lot to deal with".

No sentence of that form is appropriate as an answer to the question:

"How many presidents will there be in 2045?" And given any sentence of that form, it will always be inappropriate to say "that is false: there will not be a unique U.S. president in 2045." In other words, any sentence having the same form as (**) will *behave* just like sentences containing definite descriptions. Recall what we said a moment ago. Even when (ii) is functioning attributively, it is still inappropriate to respond to it by saying: "that is false: there will not be a unique U.S. president in 2045." And it would be inappropriate to use (ii) as a response to the question: "How many presidents will there be in 2045?" So, in those respects, (ii) *acts* just like the sentence "Smith will have a lot to deal with" or "Jones will have a lot to deal with". So if the referentialist is right, then semantically (ii) has the same form as (*); and it is no wonder that (ii) cannot be used as a means of *asserting* how many U.S. presidents there will be in 2045.

So, if the referentialist is right, then everything falls squarely into place. First, it becomes clear why (ii), and sentences of its ilk, *communicate* propositions about concepts, and not about particular individuals; it becomes clear, in other words, why they function attributively. At the same time, if the referentialist is right, it becomes clear why (ii) behaves differently from

(iii) In 2045, exactly one person x will be U.S. president, and x will have a lot to deal with.

For if the referentialist is right, then (iii) encodes something having the form:

(*) so and so will have a lot to deal with,

and this will be known to anyone who knows the semantics of (ii). So to any such person, (ii) would not register as synonymous with (iii). Thus, the referentialist's position neatly explains all the phenomena connected to the referential-attributive distinction; indeed, such an explanation is built right into that position.

Bibliography

Meaning and Truth. 1991. Jay L. Garfield and Murray Kitely (eds.) St. Paul, Minnesota: Paragon House

The Philosophy of Language. (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.

Bach, Kent. 1984. *Thought and Reference.* Oxford: Clarendon.

Barwise, J. and Cooper R. "Generalized Quantifiers and Natural Language". *Linguistics and Philosophy* 4 (1981), 159-219.

Barwise, Jon and Perry, John.1990/1983. "Semantic Innocence and Uncompromising Situations". In *The Philosophy of Language.* (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.
-- *Situations and Attitudes.* 1999. Palo Alto: CSLI Publications.

Blackburn, Simon. 1984. *Spreading the Word.* Oxford: Clarendon.

Brandom, Robert. 1998. *Making it Explicit.* Cambridge: Harvard University Press.
Burge, Tyler. 1979. "Individualism and the Mental". In French, Uehling, and Wettstein (eds.) *Midwest Studies in Philosophy* IV, Minneapolis: University of Minnesota Press, pp. 73-121.
--1982. "Other Bodies". In Woodfield (ed) *Thought and Object.* Oxford: Oxford University Press.
Carnap, Rudolph. 1934. *The Logical Syntax of Language.*
--1947. *Meaning and Necessity.* Chicago: University of Chicago Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge: M.I.T. Press.
--1975. *Reflections on Language.* Reprinted in *On Language* (1998). New York: St. Martin's Press.

Coulter, Jeff. 1983. *Rethinking Cognitive Theory.* New York: St. Martin's Press.

Church, Alonzo. "Carnap's Introduction to Semantics". *The Philosophical Review.* Vol. LII. No. 3. 1943. 298-304.

--1956. *Introduction to Mathematical Logic.* Princeton: Princeton University Press.

Davidson, Donald. 1984. "True to the Facts". In *Essays on Truth and Interpretation.* Oxford: Oxford University Press.

--1984b/1965. "Theories of Meaning and Learnable Languages".

--1990/1967. "Truth and Meaning". In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Donnellan, Keith. 1990/1966. "Reference and Definite Descriptions." In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Dummett, Michael 1981. *Frege: Philosophy of Language.* Second Edition Cambridge: Harvard University Press.

Evans, Gareth. 1990/1974. "The Causal Theory of Names". In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

-- 1980. "Pronouns". *Linguistic Inquiry* 11: 337-362.

-- 1982. *The Varieties of Reference.* Edited by John MacDowell. Clarendon: Oxford University Press.

Fodor, Jerry. 1975. *The Language of Thought.* New York: Thomas Y. Crowell

--1981a. "Introduction". In *Representations.* Cambridge: The MIT Press.

--1981b. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology". In *Representations.* Cambridge: The MIT Press.

--1987. *Psychosemantics.* Cambridge: The MIT Press.

-- 1990a "Fodor's Guide to Mental Representation". In *A Theory of*

Content and Other Essays. Cambridge. The M.I.T. Press.

-- 1990b. "A Theory of Content I". In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.

--1990c. "A Theory of Content II". In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.

--1990d. "Substitution Arguments and the Individuation of Beliefs". In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.

--1994a. *The Elm and the Expert: Mentalese and Its Semantics.* Cambridge: M.I.T. Press.

--1994. "Fodor". In *A Companion to the Philosophy of Mind.*(Samuel Guttenplan ed.). Cambridge: Blackwell.

--1998. *Concepts.* Clarendon Press: Oxford.

-- 2001. *The Mind Doesn't Work that Way.* Cambridge: M.I.T. Press.

Frege, Gottlob. 1997/1879. "Begriffsschrift". In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.

1997/1891. "On Function and Concept." In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.

--1990/1982a. "On Sense and Nominatum." In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.

-- 1953/1884. *The Foundations of Arithmetic.* Translated by J.L. Austin. Oxford: Basil Blackwell. Reprinted in 1980 by Northwestern University Press.

-- 1997/1892b. "On Concept and Object." In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.

Geach, Peter. 1962. *Reference and Generality.* Ithica: Cornell University Press.

Gödel, Kurt. 1944. "Russell's Mathematical Logic." In *The Philosophy of Bertrand Russell.* Edited by Paul Arthur Schilpp. Evanston, Illinois: Northwestern University Press.

Jackson, Frank; Pettit Philip; Smith, Michael. 2004a. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004b. "In Defense of Folk Psychology." In Jackson, Frank; Pettit

Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004c. "Causation in the Philosophy of Mind". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- "Functionalism and Broad Content". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004e. "Program Explanation: A General Perspective". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004f. "Structural Explanation in Social Theory". In Jackson, Frank; Pettit Philip; Smith, Michael. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

-- 2004g. "In Defense of Explanatory Ecumenism". In Jackson, Frank; Pettit Philip; Smith, Michael. 2004e. *Mind, Morality, and Explanation.* Oxford: Oxford University Press.

Kaplan, David 1964. *Foundations of Intensional Logic.* PhD dissertation. University of California, Los Angeles.

--1990/1970. "Dthat". In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Kaplan, David. 1978. "On the Logic of Demonstratives". *Journal of Philosophical Logic* 8 (1978): 81-98.

-- 1989. 'Demonstratives'. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

-- 1989b. "Afterthoughts". In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

Kripke, Saul. 1999/1971. "Identity and Necessity". In *Metaphysics: An Anthology.* Edited by Jaegwon Kim and Ernest Sosa. Oxford: Blackwell.

--1972. *Naming and Necessity.* Cambridge: Harvard University Press.

--1974. The Locke Lectures (unpublished).

--1991/1977 ‘Speakers Reference and Semantic Reference’. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

-- "A Puzzle about Belief". 1979. In *Meaning and Use*, ed. A. Margalit. Dordrecht: Reidel.

Lewis, Clarence. 1946. *An Analysis of Knowledge and Valuation.* Cambridge: Harvard University Press.

Loar, Brian (1988) "Social Content and Psychological Content" in Grimm and Merrill (eds.) *Contents of Thought* University of Arizona Press.

Lewis, David. 1984. *On the Plurality of Worlds.* Oxford: Blackwell.

Lyons, John. 1977. *Semantics.* Cambridge: Cambridge University Press.

McCawley, James D. 1981. *Everything that Linguists Have Always Wanted to Know about Logic.* Chicago: University of Chicago Press.

--1998. *The Syntactic Phenomena of English.* Chicago: University of Chicago Press.

McGinn, Colin. 1988. *Mental Content.* Oxford: Oxford University Press.

Mill, John Stuart. 1991. "Of Names and Propositions". In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Montague, Richard. 1974. "The Proper Treatment of Quantification in Ordinary English". In *Formal Philosophy.* Edited by Richard Thomason. New Haven: Yale University Press.

Neale, Stephen. 1990. *Descriptions.* Cambridge: The M.I.T. Press.

-- 1995. "The Philosophical Significance of Church's Slingshot." *Mind*, New Series, Vol. 104, No. 416 (Oct., 1995), 761-825.

--2001. *Facing the Facts.* Oxford: Clarendon Press.

Olson, Ken. 1987. *An Essay on Facts.* Palo Alto: CSLI Publications.

Putnam, Hilary. 1962. "The analytic and the synthetic" In Herbert Feigl and Grover Maxwell, eds., *Minnesota Studies in the Philosophy of Science*, III, Minneapolis: University of Minnesota Press.

--1975. "The Meaning of 'Meaning'". In *Collected Papers: Volume II.* Cambridge: Harvard University Press.

Pylyshin, Zenon W. 1984. *Computation and Cognition.* Cambridge: The M.I.T. Press.

Quine, Willard van Orman. 1990/1951. "Two Dogmas of Empiricism." In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

--1953. "Reference and Modality." In *From a Logical Point of View.* Cambridge: Harvard University Press.

--1960. *Word and Object.* Cambridge: The M.I.T. Press.

-- 1966a. "Three Grades of Modal Involvement." In *The Ways of Paradox.* Cambridge: Harvard University Press.

--1966b. "Vagaries of Definition". In *The Ways of Paradox.* Cambridge: Harvard University Press.

-- 1966c. "Quantifiers and Propositional Attitudes". In *The Ways of Paradox.* Cambridge: Harvard University Press.

-- 1968. *Philosophy of Logic.* Cambridge: Harvard University Press.

Robinson, Daniel Nicholas. 2002. *Praise and Blame.* Princeton: Princeton University Press.

Russell, Bertrand. 1903. *Principles of Mathematics.* Cambridge: Cambridge University Press.

--1990/1905. "On Denoting". In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.

-- 1957/1910. "Knowledge by Acquaintance and Knowledge by

Description". In *Mysticism and Logic.* Garden City, New Jersay: Doubleday.

-- 1920. *Introduction to Mathematical Philosophy.* London: George Allen Unwin.

--1921. *The Analysis of Mind.* London: George Allen Unwin.

-- 1948. *Human Knowledge: Its Scope and Limits.* London: George Allen Unwin.

Salmon, Nathan. 1986. *Frege's Puzzle.* Cambridge: M.I.T. Press.

-- 2002. "Mythical Objects". In *Meaning and Truth.* New York: Seven Bridges Press.

Stalnaker, Robert. 1991. "Propositions". In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Soames, Scott. 2001. *Beyond Rigidity.* Princeton: Princeton University Press.

Strawson, Peter. 1990/1950. 'On Referring' in A.P. Martinich (ed) *The Philosophy of Language.* Oxford: Oxford University Press.

Stroud, Barry. 2001. *The Quest for Reality.* Oxford: Oxford University Press.

Tughendat. Ernst. 1970. "The Meaning of 'Bedeutung' in Frege". In *Analysis*., (1970), pp. 177-189.

Weisler, Stephen. 1991. "An Overview of Montague Semantics." *Meaning and Truth.* 1991. Jay L. Garfield and Murray Kitely (eds.) St. Paul, Minnesota: Paragon House

Whitehead, Alfred North and Russell, Bertrand. (1997/1912). *Principia Mathematica (to *56).* Cambridge: Cambridge University Press.

Wittgenstein, Ludwig. 1958. *The Philosophical Investigations.* Translated by G.E.M. Anscombe. Oxford: Basil Blackwell.

--1983. *Remarks on the Foundations of Mathematics.* Edited by G.H. von Wright, Rush Rhees, and G.E.M. Anscombe. Cambridge: M.I.T. Press.

---

[1] Davidson 1984.

[2] Carnap (1947) makes this point.

[3] See Davidson 1984, Evnine.

[4] Russell (1903) said that "P is a proposition" is equivalent to "P entails P". He is right. But that probably doesn't qualify as an *analysis* of the concept of a proposition. (I don't think Russell was trying to give an analysis.)

---

[1] Frege (1982a) actually gives two quite distinct arguments for the thesis that sentences refer to truth-values. For reasons of space, I will consider only the second (the Slingshot). If cogent, the first argument – the one I don't give – shows that sentences *do* refer to truth-values. If cogent, the second one – the famous one (the "Slingshot") – only shows that *if* sentences refer to anything, it is to truth-values.

[2] Frege 1892a.

[3] This is probably no *analysis* of the concept of reference – it is merely a platitude. But all we need in this context is a platitude.

[4] It should be pointed out, by the way, that the sense in which *the inventor of bifocals has phi* implies this is not comparable to the sense in which "1+1=2" implies that arithmetic is incomplete. The entailment from "1+1=2" to "arithmetic is incomplete" is mediated by a great deal of logic; it is not merely *semantic*. By contrast, the entailment from *the inventor of bifocals has phi* to *somebody x invented bifocals, and x has phi* doesn't involve the mediation of any logic; it is purely *semantic*. Merely in virtue of understanding the first, you know that it involves the truth of the second: nothing more than semantic competence is needed. Anyone who understands – who grasps the semantics of – "the inventor of bifocals", knows that if there

are five (equally contextually salient) inventors of bifocals, or zero inventors of bifocals, then something is very wrong with "the inventor of bifocals smokes". Nothing more than *semantic* competence is needed to see this; no logical competence (except such as is required for semantic competence) is involved. So if Anderson's semantics is right, it is a matter of *semantics*, and not of logic, that "the inventor of bifocals snores" says something about the property of inventing bifocals; it is a matter of semantics that, in virtue of being of the form "…the inventor of bifocals…", a sentence says something about that property (in fact, it encodes a proposition that has that property as a constituent). So Anderson's semantics seems to lead to the view that "the inventor of bifocals" refers, at least in part, to that property (and thus, by implication, doesn't *wholly* co-refer with "the first post-master general", since the latter expression doesn't refer, even in part, to the property of inventing bifocals).

[5] See Russell 1920: 167-168.

[6] In private correspondence.

[7] I am slightly adapting Church's argument.

[8] The asterices are meant to be quasi-quotes. My computer doesn't do quasi-quotes.

[9] Davidson 1967. The technique used in his argument is really due to Church 1943.

[10] Russell 1920 (165).

[11] It is a delicate question whether that pair comprises the *sentences* or propositions. For reasons I will argue later, I think it is pretty clear that it comprises propositions. But in the present context, this is not important.

[12] This term is borrowed from Neale (2001). It is short for: Principle of Substitution of Logical Equivalents.

[13] See Perry's "Evading the Slingshot" in Perry (2001). See also Neale (1994) and Neale (2001).

[14] Some authors distinguish among "intensional" and "non-extensional" and "hyper-intensional". For our purposes it is enough that we distinguish between contexts that are extensional and those that aren't.

[15] I must make it clear that *in my view*, if the underlined clauses are to refer to different propositions, the definite descriptions must be read as quantifiers. The first underlined clause must be taken to mean something like: *the proposition that somebody uniquely invented bifocals and that individual*

snored (with the same *mutatis mutandis* holding of the second clause). And, of course, thus read, the existential quantifiers would have to be given narrow scope to ensure a difference in truth-value.

In my view, *if* the definite descriptions are really functioning as singular terms, then they both denote the proposition *Franklin snored.*

But, once again, I am here operating within the kind of Fregean framework that the Slingshot depends on. I am *granting* both that the two clauses denote different propositions *and* that the one underlined clause is what results when a referring term in the other clause is replaced with a co-referring term. Again, I don't think that it is coherent to make both these assumptions simultaneously. I am allowing that it is, for the sake of argument, in order to give the proponent of the Slingshot as much leeway as possible.

[16] See, for example, Recanati (2000: 3-4).

[17] See my disclaimer in the second to last footnote.

[18] Frege (1892: 193).

[19] Actually, once we've clarified our views on reference, we will find that there is an internal relation between the principle of compositionality and the concept of reference: the former follows from the latter, and – not merely for psychological, but purely logical reasons - any language worthy of the name must obey compositionality:

[20] In each of (1)-(4), the emphasis is Taylor's.

[21] (1) collapses into (3). Consider:

(a) "John believes <u>that four is greater than three</u>":
(b) "John believes <u>that the number of stomachs had by an average cow is greater than three</u>".

Why do (a) and (b) differ in truth-value? Because the underlined clause in (a) refers to a different proposition from the underlined clause in (b). So the reason that intensional contexts are not always truth-functional is that, in such contexts, intersubstituting co-referring terms doesn't preserve reference. So (1) is redundant; (3) does its job.

[22] Obviously we are talking about the complementizer "that", not the demonstrative.

[23] It occurs to me that Russell's "Grey's Elegy" argument (which I do not claim to understand) is similar to this.

[24] And, I believe, it is demonstrably incoherent.

[25] See Carnap 1934. These few hold this view. But Stalnaker 1991 is an exception. So, perhaps, is David Lewis.

[26] Davidson 1967.

[27] 1964 (77).

[28] Except, of course, in the special case where the thing *to which* a truth-value is being assigned is itself a truth-value. Consider the open-proposition *x is a unique truth-value T such that* snow is white *has T*.

[29] I myself believe, and will argue, that concepts *are* properties. But this cannot be taken for granted since there *seem* to be cases of distinct concepts that correspond to the same property – for example, the concepts *success of one* and *unique even prime* are obviously distinct. But a thing cannot possibly have the one property unless it has the other, raising the question whether we are dealing with different properties.

[30] Oftentimes, the term "logical equivalence" is confined to expressions that are "formally" equivalent, e.g. "P" and "either P or (Q and not Q)". Here we are using that term more broadly. We will describe as "logically equivalent" propositions like *triangles have three sides* and *1+1=2*, and will extend this policy to the corresponding sentence-tokens.

Incidentally, I think that the notion of so-called *formal* equivalence is much more contextual than is generally thought; and the belief that it has an absolute meaning underlies some (in my view) quite erroneous beliefs about the mind.

[31] I think that (ii) is not *quite* right. I think that, strictly speaking, the rule is this: *if* at context C [where context includes time], O is a unique *salient* U.S. President, *then* a token in C of "the current U.S. President" refers to O. We will deal with this sort of nicety in th second half of the book. It doesn't have any importance in this context.

[32] There is more to it than that. (iv) gives a "macro-rule"; it doesn't give the full derivation tree. But it gives what is, in this context, the relevant

component of the relevant semantic rule.

[33] This is a very approximate way of stating the rule for "here". Suppose Brown is in Paris, and says "Smith is here". Brown might mean: Smith is in *this room* or Smith is in *this country* or even Smith is in *this hemisphere*. What is denoted by a token of "here" is some conversationally relevant area that *includes* but is not necessarily *confined to* the specific place where the tokening occurs. The area might include only the room *in which* the tokening occurs – but it might also include the city, or the county, or the country, or the continent in question. So the right rule for "here" is something like this. If a sentence of the form "…here…" is tokened in place P, and P* is the conversationally relevant area including P, then that token is true exactly if P* has…x… So if Brown is in some specific place, and Paris is the relevant area that includes that specific Place, then when Brown says "Smith is here", Brown's token is true exactly if Smith is in Paris.

Exactly similar remarks apply to the term "now". "now" may refer to a certain minute, a certain hour, a certain century, a certain geologic area. So, strictly speaking, the right semantic rule for "now" is not the one I give above, but is rather something like this. If a sentence of the form "…now…" is tokened at instant t, and T is the conversationally relevant a period of time that encompasses t, then if a sentence of the form "…now…" is tokened, that token is true exactly if T has…x…So suppose, at instant t, Jones says "now most nations are democracies", and the period of time following 1770 is the relevant stretch of time encompassing t, then Jones' utterance is true exactly if, during the period following 1770, most nations are democracies.

But these niceties relating to the semantic rule for "here" are not relevant in this context; and, in any case, it is very easy to amend our discussion so as to allow for them.

[34] I am innocuously modifying his argument

[35] See Anderson 1998, page 147.

[36] See Anderson 1998, page 147, footnote 49.

[37] See Anderson 1975, who uses this point as a weapon against Russell's Theory of Descriptions.

[38] See Anderson 1998, page 147, footnote 49.

[39] Except Gödel's, but that one is still null and void, though for different reasons.

[40] Let us momentarily idealize away from the type-theoretic problems relating to the concept of *everything* and, in connection with that, of a universal class.

[41] In actuality, I think that it probably takes for its arguments n-tuples consisting of worlds, times, possibly places, and classes.

[42] Strawson 1950.

[43] Given what Kripke said, this is almost certainly false. But Frege's argument requires that Kripke be wrong. So for the sake of argument, let us grant this to Frege.

[44] Russell 1920.

[45] But not decisive. A Fregean would say that definite descriptions are singular terms, but that "the inventor of bifocals snored, but the first post-master general did not snore". *Prima facie* this point is compelling. I myself think that this line of thought breaks down when we consider the distinction between type-semantics and token-semantics and between semantics and pre-semantics. And I myself think that, if tokens of that sentence really were non-self-contradictory, then the definite descriptions would be quantifiers, not singular terms. But obviously these points are less than self-evident. So the Fregean response to Russell's point has some initial force.

[46] Let

(d) "Mozart wrote music and Bach also wrote music"

be a sentence-token. Even in (d), the occurrences of "Mozart wrote music" and "Bach also wrote music" do not by themselves have assertoric force. What has force is the *whole* of (d). Somebody uttering (d) would be *committed* to the truth of the component sentences. But, as Geach made clear, he would not have *directly* asserted them; he would have *directly* asserted only the conjunction. See Geach 1972: 259. See also Church 1956: 22-23.

[47] There are some cases where sentences seem to have "mixed" forces, e.g. "if he comes home, tell him to call me". Here it might seem as though the *component* "tell him to call me" has force all on its own. I don't think this is the case. See Dummett (1973: 163). A whole article devoted to that particular topic is__ , by the present author: there it is argued that it is impossible that components of a sentence (or, more accurately, sentence-token) can have any kind of force; and apparent problems – like tokens of "if

he comes, should I tell Fred?" – are dealt with.

[48] I myself think it is possible to come up with a *bona fide* analysis of reference. But a defense of that conception would itself involve an entire paper. But here I will give an outline. First of all, for reasons we saw a moment ago, reference is probably a property of tokens, not types. So what refers is (if anything) a token of "the inventor of bifocals", not the corresponding type. In effect, we earlier saw some reason to believe that *if* a token of "the inventor of bifocals" is really a singular term that picks out Franklin, then a token of *the inventor of bifocals has phi* encodes the proposition *Franklin has phi*: so the proposition in question has *Franklin as a constituent.* And we saw some reason to believe that if some kind of *concept* or *sense* makes it into the proposition meant by a token of *the inventor of bifocals has phi*, then the token of "the inventor of bifocals" ends up being a quantifier, not a singular term. I think that, for extensions of those reasons, *if* E is a referring term that refers to x, then sentence-tokens of the form "… E…" encode propositions that have x *as a constituent.* If what makes it into proposition isn't x itself, and is instead some sense or concept, then E is a quantifier, as opposed to a singular term that refers to x. So I believe that if E refers to x, then sentence-tokens of the form "…E…" encode propositions that have x as a constituent.

I also believe that the converse holds. If sentence-tokens of that form encode propositions that have x as a constituent, then E refers to x. Suppose that any sentence-token of the form *Argo has phi* ipso facto encodes a proposition means that Margaret Thatcher has phi (but suppose that, otherwise, English semantic rules are unchanged). In that case, surely "Argo" (or tokens thereof) would refer to Margaret Thatcher (or, at the very least, E would have a component that thus referred).

So it seems that E refers to x exactly if, in virtue of having the form "…E…", a sentence-token encodes a proposition that has x as a *constituent.*

To sum up: E refers to x exactly if, in virtue of having the form "… E…", a sentence-token encodes a proposition that has x as a *constituent.*

So contrary to what Church (1964) held, I think that "denotation" is *not* a primitive concept. (See also Anderson 1998, page 152, footnote 55.) But as I said a moment ago, in order for my case for this to be water-tight, I'd have to deal more thoroughly with the Fregean view of reference – namely, that E can refer to x without x's being an actual constituent of the proposition meant by "…E…". (Significantly, Church was a hardcore Fregean.) I've

already given a general argument against the Fregean position. But, of course, a thorough argument would require a paper (or book – or series of books) unto itself.

[49] We should consider a sentence-token that might seem *not* to conform to this analysis:

(a) "Mozart wrote music or grass is green".

The occurrence of "Mozart wrote music" doesn't have any kind of sentential force. But, on the face of it, it seems that (a) doesn't attribute *any* property to the proposition *Mozart wrote music.* So here it appears we have a counter-example to our analysis.

But on closer inspection, this appearance proves to be dubious, at best. As a matter of semantics, (a) is true exactly if *either* of *Mozart wrote music* or *grass is green* is true. So (a) can be seen as attributing to the former the property of being a thing x such that either x or *grass is green* is true. And it can be seen as attributing a similar property to *grass is green*.

Here we would do well to consider a possible objection.

A token of "Plato snores" is true iff the proposition *that Plato snores is true* is true. But a token of "Plato snores" is about Plato and snoring; it is not (except by implication) about a proposition. Of course, *as a matter of logic*, "Plato snores" is true iff a certain proposition is true. But that is not enough to warrant the view that "Plato snores" is itself *about* that proposition. *As a matter of logic*, Plato snores" is true iff the proposition *either arithmetic is complete and there is a way to square the circle or Plato snores* is true. But "Plato snores" is not about arithmetic or about circles. (In any case, at the level of *semantics*, "Plato snores" is not about arithmetic or circles. One might say that *logically* it is about such things, since it entails them. But in that case *every* sentence would be about arithmetic. And, in any case, here we are concerned with the semantic, not the broadly logical, properties of expressions.) So semantic facts cannot be read of logical equivalences.

Now (a) is indeed true exactly if either of *grass is green* or *Mozart wrote music* is true. But that is a far cry from saying that it attributes *truth* to either of those propositions. After all, (a) is true

exactly if the proposition:

[either arithmetic is complete and there are finitely many primes] or [either *grass is green* or *Mozart wrote music*]

is true. But (a) is surely not (at least not in a strict semantic sense – we can set aside the broad logical sense) about prime numbers or the completeness of arithmetic.

In my view, this objection answers itself. The sense in which (a) is about the propositions *Mozart wrote music* and *grass is green* is obviously not remotely comparable to the sense (if there be any) in which it is about arithmetic or the number of primes. No one could speak English without knowing that, for (a) to be true, at least one of those two propositions must be true. One doesn't have to be a logician or mathematician to see this dependency: a basic grasp of English semantics is enough. So by itself, unsupplemented by so much as a whit of logico-mathematical acumen, a basic understanding of English is enough (more than enough) to see that, for (a) to be true, then either of those two propositions must be true. Since nothing more than *semantic* competence is needed to see this – since nothing more than a grasp of semantic rules is involved (no logico-mathematical dexterity is needed…except such as is a prerequisite to grasping semantic rules) – it is not unreasonable to say that, at the strictly semantic level, (a) *is* about those propositions. Anyone who grasps the semantics of (a) – we can set aside a grasp of logico-mathematical theorems – knows that if *either* of *Mozart wrote music* or *grass is green* is true, then (a) is true. So it is not unreasonable to say that, at the semantic level, (a) attributes to *Mozart wrote music* the property of being such that either it or *grass is green* is true (the same thing *mutandis* holding of *grass is green*). I think it is very misleading to say that (a) is *about* those propositions only in the recherché sense in which (a) is *about* prime numbers or the properties of the real number system. (a) is obviously *about* those propositions in a strictly and narrowly semantic sense. And anyone who failed to see that the truth of either *grass is green* or *Mozart wrote music* suffices for the truth of (a) would *ipso facto* fail to grasp the *semantics* of that sentence-token: he would fail to grasp *not* exotic facts about its logico-mathematical liasons, but about its linguistic

meaning. And this suggests that, *at the level of linguistic meaning* – at the narrowly semantic level – (a) attributes the aforementioned properties to those propositions. So I think that the objector's point is spurious; and I don't think that, in (a), we have anything remotely resembling a *clear* counter-example to our thesis that forceless sentence-tokens refer to propositions.

[50] See Black 1954: 235-236. Black talks about the property of being meaningful, not of being grammatical. And his argument is meant to refute the thesis that sentences refer to *truth-values*. See the last footnote, where I discuss the merits and demerits of Black's argument.

[51] See Geach (1972: 259).

[52] As I discussed in a previous footnote, there are a few dubious exceptions to this, e.g. "if he comes home, *tell him to call me*." See Dummett (1973: 163).

[53] See Russell 1981, Neale 2001.

[54] Although even this is doubtful, when we take into account facts about intonation, capitalization punctuation, and the like.

[55] See Russell 1919.

[56] Kaplan (1970).

[57] See the correspondence between Frege and Russell in the Salmon-Soames anthology.

[58] Review of *Frege on Functions* by Max Black. Journal of Symbolic Logic 1956 (21): 201-202.

[59] Actually, I will argue that "John" picks out a third-level function. But we can ignore that here.

[60] See Dummett 1973 (192-193) for similar remarks.

[61] Frege 1892a.

[62] Mill 1991 (10).

[63] Church 1956.

[64] Here, of course, people will bring up Donnellan's (1990/1966) point that, even when "the man drinking the martini" is functioning as a *referring* expression (not a quantifier), "the man drinking the martini smokes" makes a true statement (or, better, can be *used* to make a true statement) even if there is no man drinking a martini. And this fact might call into question the accuracy of my statement that (REF) captures the semantics

of "the phi" for referentialists.

This is well-trodden territory. It is pretty clear that *if* "the man drinking a martini" is a referring expression, *then* at the level of literal meaning "the man drinking the martini smokes" is abortive (neither true nor false) if there is no one drinking a martini. Granted, that sentence may say something true. But that is very easily explained as a pragmatic epiphenomenon. And it would be an egregious violation of Grice's "modified Occam's razor" (never multiply meanings without necessity) to say that *semantically* "the man drinking the martini" actually referred to (say) the female transvestite drinking the ginger ale.

[65] I am idealizing away from the pragmatic epiphenomena mentioned in the last footnote. I don't think they have any relevance here.

[66] Let us consider a case where *by our stipulation* we have a referring term whose referent we do not know. The problems that arise in connection with definite descriptions arise in connection with this term. We will find that there are no mysteries as to how to deal with them.

If there is an entity x uniquely having property phi, let "ze phi" denotes x. If there is no such entity, then "ze phi" denotes nothing.

If the referentialist account of definite descriptions is right, then "ze finance minister of Zimbabwe in exactly twelve years" is synonymous with "the unique person who will be finance minister of Zimbabwe in twelve years".

You and I certainly do not know who "Ze finance minister of Zimbabwe in exactly twelve years" names.

Consider the sentence:

(ZF) "Ze finance minister of Zimbabwe in exactly twelve years is over 6 ft tall".

Do you and I understand (ZF) or not?

Strictly speaking, unless we know who "ze finance minister of Zimbabwe in exactly twelve years" refers to, we don't know what the literal meaning of (ZF) is.

But even if we don't know who the zefinite description refers to, we still know a great deal about what (ZF) means. We know it will be true exactly if:

(*) somebody is uniquely a finance minister of Zimbabwe in twelve years, and any such person is over 6ft tall.

So "ze finance minister of Zimbabwe in exactly twelve years is over 6ft tall" will *communicate* (*) to us, even if we have no idea who the zefinite description refers to.

Notice that (*) is exactly the proposition that, according to Russell, is encoded in:

(RF) "The finance minister of Zimbabwe in exactly twelve years is over 6 ft tall".

(ZF) has exactly the same cognitive value as (RF).

Supposedly, Russell's theory is supported by facts about the cognitive values of sentences containing definite descriptions.

In actuality, such facts give *no* support to Russell's theory. For those very facts equally support the view that definite descriptions are zefinite descriptions – that "the inventor of bifocals" is synonymous with "ze inventor of bifocals". Thus, those facts no more support Russell's position than they support the referentialist's. This means that those facts are evidentially neutral.

[67] 1982.

[68] See Kripke 1972, Evans 1990/1973.

[69] See Kripke 1971, 1972.

[70] See Kripke 1971, 1972.

[71] I put RING in brackets because, in L, it is not the English word "ring", but the *actual* ringing of some particular doorbell that occurs in that sentence-token.

[72] See Dummett 1973, and Tughendat 1970, for similar remarks.

[73] I put RING in brackets because, in L, it is not the English word "ring", but the *actual* ringing of some particular doorbell that occurs in that sentence-token.

[74] This extra-semantic rule is phenomenally ungainly and useless. But it does apparently qualify as a genuine semantic rule.

[75] Blackburn (1984) argued that Tarski's paper "The Concept of

Truth in Formal Languages" disjunctives the concept of truth: there is trueinEnglish, trueinUrdu, and so on. This is perhaps not quite on point. But I thought it worth mentioning.

[76] Fodor 1998.

[77] See Fodor 1990a and 1990b for an attempt to deal with what are essentially these problems. Fodor is dealing not with the causal theory of reference, but with the causal theory of *conception*: for x to be a mental representation of y simply *is* for x to have a certain causal relation to y. The causal theory of conception is exactly analogous to the causal theory of reference. The problems that face the one are exact analogues of the problems that face the other. Fodor's two papers, though ingenious, show how implausible those theories are.

[78] Kaplan 1989b (604-607) makes some remarks that are similar to these.

[79] Kaplan 1989b (604-607) makes some remarks that are similar to these.

[80] Kaplan 1989b (568)

[81] 1964 (77).

[82] See Dummett 1973 (192-193) for similar remarks.

[83] Church1944 (129-130).

[84] On the Fregean view, "1+1=2" and "triangles have three sides" do indeed co-refer. Indeed, on the Fregean view, they co-refer, as a matter of logical necessity.

At first this seems to be an advantage over my view.

It is not.

When two *sentences* are said to be "logically equivalent", S-equivalence is not what is meant. If you ask a competent person, "what is it for two sentences to be logically equivalent?", he absolutely will not say: "it is for the open proposition that a thing must satisfy to be picked out by the one expression to entail, and be entailed by, the open proposition that a thing must satisfy to be picked out by the other." What he *will* say is: it is for the proposition meant by the one to entail, and be entailed by, the proposition meant by the other. So the kind of "logical equivalence" that people have in mind when they talk of logically equivalent *sentences* is not the kind of "logical equivalence" that people have in mind when they talk about expressions like "Plato" and "the unique x such that 1+1=2 and x=Plato". When it is said that two *sentences* are "logically equivalent", S-equivalence is

not meant; so the kind of "logical equivalence" that guarantees co-reference is *not* what is meant. Some completely different kind of logical equivalence is meant. And this other kind of logical equivalence has very little to do with co-reference. So to think that "1+1=2" and "triangles have three sides" must co-refer because they are "logically equivalent" is to make the false and confused assumption that those two sentences are logically equivalent in the sense in which "Plato" and "the unique x such that 1+1=2 and x=Plato" are logically equivalent. We have two very different notions of logical equivalence here, not one. The one kind of logical equivalence involves co-reference. But that leaves it totally open whether the other kind involves co-reference.

Let us close the argument. It is often thought that 1+1=2" and "triangles have three sides" must co-refer. Frege thought this; so do all proponents of Slingshot arguments. The reasoning behind this view is:

Those two sentences "logically equivalent"; and "logically equivalent expressions" – e.g. "Plato" and "the unique x such that 1+1=2 and x=Plato" – must co-refer.

But that view involves a complete failure to distinguish two kinds of "logical equivalence". And the view that a correct semantic theory must make logically equivalent sentences co-refer is based on that very confusion. So, therefore, is the view that Frege's analysis has an advantage over mine, in that his view makes "logically equivalent" sentences co-refer while mine does not. The idea that that is an advantage involves a confusion of two very different meanings of the term "logically equivalent".

[85] Some historical background may put this into context. There are two very different conceptions of what linguistic meaning is. Some see sentence-meaning, and expression-meaning generally, in non-psychological terms. The meaning of "some" is a function from classes to truth-values; the meaning of "Plato" is an individual, or a function from contexts of utterance to that individual. The meaning of an expression is always a non-psychological entity, except in the special case where the expression in question denotes a mental content.

Let us refer to advocates of this view as "semantic objectivists".

Others see sentence-meaning in psychological terms. The meaning

of a sentence-token is what the *speaker* means by it. Advocates of this view are sometimes referred to as "speech-act theorists". When you perform a speech-act – for example, when you make an assertion or ask a question – your words mean what *you* mean.

Of course, speech-act theorists grant that if someone says "I am cold" but means *I am hot*, his words don't mean what *he* means. I am simply outlining it. My statement of his position is meant only to be a bare approximation.

Speech-act theorists tend to embrace a semantic nihilism. They tend to think that the notion of *the* meaning of an expression-type is bankrupt: when we talk about *the* meaning of "snow is white", that is really just an oblique and misleading way of talking about what *people* mean when they produce those sounds or sounds relevantly similar to them; they think that talk of "meaning", in the platonic sense, is really just a confused way of talking about "meaning", in the psychological sense.[85]

I don't think that speech-act theory requires any kind of anti-Platonism. (In fact, I think it requires a kind of Platonism.) But, as a matter of historical fact, speech act theorists tend to be anti-Platonistic; they tend to say that meanings, in the non-psychological sense, don't exist.

So speech-act theorists tend to say, in effect, that semantics doesn't exist. Semantics is typically thought of as an attempt to associate the platonic entities – the right functions, the right sets of worlds, the right properties – with expressions. The speech act theorist tends to reject this approach. So he rejects the discipline of semantics, at least as most semanticists conceive it.

For some reason, semantic-objectivists talk very little about force. But speech-act theorists talk a great deal about it. Since the only people who talk about force are the ones who reject the very notion of semantics, the impression has been created that force has no place in discussions of semantics.

This is plainly a mistake. Even those who have an extremely hard-line objectivist conception of meaning, such as the one advocated in this work, must deal with the concept of force. And they *can* deal with it. Psychology is no more needed to deal with the concept of force than it is needed to deal with the concept of variable-binding.

(I think psychology does have an important connection to semantics. But the connection is that mental entities *hook up* platonic entities

to expressions, and thus *give* certain meanings to those expressions. The connection is *not* that mental entities *constitute* the meanings of expressions. I don't think there is any real opposition between semantic objectivism and speech-act theory. One is focusing on the psychological mechanisms that assign meanings to expressions. The other is focusing on the meanings themselves.)

[86] Wittgenstein 1958, Dummett 1973.

[87] See Dummett 1973 (192-193) for similar remarks.

[88] 1999, 1990/1983.

[89] Kripke 1991/1977.

[90] Of course, "that grass is green" is not, at least not phonetically, identical with "grass is green". And (though I think this is implausible) one might suppose that dropping the "that" induces a change in reference: previously a proposition was referred to; now a truth-value is referred to.

But then we are just stuck, once again, with the view that "grass is green" refers to a truth-value in (p) and either refers to nothing, or refers to a proposition, in (n). And this is just what we want to avoid saying.

[91] We know from Kripke that proper names of individuals, and probably of natural kinds, are semantically associated with senses. Semantically "Smith" does not have a sense. Semantically, "iron" probably does not have a sense. ("H_2O" does have a sense, but that is because it is really a compressed definite description: "the unique substance whose constituent molecules consist of two hydrogen atoms and one oxygen atom".)

Frege said that epistemic operators cause sentences to refer to the propositions that they ordinarily refer to. So

(i) "Plato smoked"

refers to a truth-value, when it occurs on its own. But it denotes its customary sense in:

(ii) "Bob believes that Plato smoked"

And (ii) refers to a truth-value when it occurs on its own; and it denotes its customary sense in:

(iii) "Fred believes that Bob believes that Plato smoked".

A consequence is that "Plato" denotes its customary sense in (ii), and the sense of its sense in (iii).

Here is a problem with that view. We know from Kripke that, at the level of semantics, "Plato" doesn't have *any* customary sense. So, if Frege is right, "Plato" doesn't denote *anything* in (iii).

Frege's theory requires that "Plato" semantically have a sense. It doesn't; so his theory is wrong.

Actually, Frege's theory requires that "Plato" be *infinitely* ambiguous in two quite distinct ways. First, it requires that it mean one thing in (i), another thing in (ii), a third thing in (iii), a fourth thing in

(iv) "Fred believes that Bob believes that Plato smoked".

So Frege's theory makes "Plato" infinitely ambiguous in that way.

But we must remember that any object can be given by infinitely many different modes of presentation or senses. There are infinitely many descriptions that single out anything. So, if Frege's theory is right, then "Plato" is finitely ambiguous in (iii), infinitely ambiguous in (iv), and so on.

If you take the position that "Plato smoked" means one thing on its own, and a different thing when it falls in the scope of an epistemic operator, then you require two things that cannot be granted. You require that "Plato" semantically have a sense. This cannot be granted, given what Kripke said.

Second, you require that "Plato" be infinitely ambiguous. Indeed, you require that it have an infinitely large set of meanings in (ii), a *second* infinitely large set of meanings in (iii), a *third* infinitely large set of meanings in (iv), and so on. This cannot be granted.

[92] Davidson 1984b/1965.

[93] Neale 2001,1994.

[94] In the Slingshot-literature, I've never seen so much as passing reference to Davidson's argument. Neale (2001, 1994) never mentions it.

[95] Chomsky 1965.

[96] Russell 1990/1905.

[97] Quine 1966c.

[98] Quine 1966c.

[99] Quine 1966c.

[100] This is what linguists tend to say about it. See McCawley 1998.

[101] I believe Russell 1905 was the first person to make this important distinction.

[102] See Robinson 2002 (135).

[103] See Barwise and Cooper (1981: 164).

[104] Russell 1903.

[105] Lyons 1977.

[106] Burge 1979 gives a similar argument, purporting to show that one's *social* environment has a role in the individuation of our thoughts similar to that which, if Putnam is right, is had by our physical environment. Kent Bach 1984 criticizes Burge's argument. I am sympathetic to Bach's point; but I am not sure how much what I am saying here is in the same vein as what Bach says.

[107] See Burge 1986.

[108] See Burge 1986.

[109] Unless, of course, one takes a Berkeleyan stance.

[110] In any case, if such propositions cannot be true, that doesn't have anything to do with the external world; it has to do with more general problems attaching to self-knowledge: Freudian problems, problems relating to the integratedness of the knower and the known, and so on.

[111] The concept of "proof" is more vague than is usually thought. I am not referring to the kinds of considerations Wittgenstein (1983) adduces. "Proof", I think, is a contextual notion. You don't prove that there are infinitely many primes. You prove that *if* such and such propositions are accepted, *then* there are infinitely many primes.

[112] Clarence Lewis (1946), Rudolf Carnap (1934), Robert Brandom (1998), and many others have said that a proposition P is the class of all its entailments or logical consequences. But they all neglected the question: If that is the case, then why are the propositions *triangles have three sides* and *1+1=2* so obviously different? That is the question to which I now would like to outline an answer: I am going to propose an amendment to the Carnap-Lewis position that is not vulnerable to this problem.

First of all, why take the Carnap-Lewis position or anything like it? Suppose that P entails Q (i.e. that Q is an analytic consequence of P). In

that case, there is surely some kind of *internal* connection between P and Q. [112] In other words, the entailment is underwritten by some kind of overlap of *content.* Part of what you are saying (though you may not know it) when you affirm P *is* Q.

For P to entail Q is really for it to be the case that [P and not Q] is self-contradictory. The self-contradiction may not be *obvious.* It may not be *obviously* self-contradictory to say: *x is a triangle but it is not the space bounded by three lines such that any two of them intersect but not all three of them intersect.* But it is a self-contradiction no less. We shouldn't put too much logical stock in a psychological phenomenon like obviousness, or the lack thereof

Now if Q were no part of the content of P, then it is hard to see how [P and not Q] would be self-contradictory. So if we are to do justice to the relation that holds between P and Q, when the former entails the second, we must see the second as being *part* of the content of the former.

But, as we've seen, it is not enough to say that a proposition is the class of its consequence. For reasons considered above, we must say: Those propositions have different "analytic profiles"; and that is why they are distinct.

So P and P* are the same proposition exactly if (i) for any proposition P**, P entails P** exactly if P* entails P** and (ii) the *analytic route* between P and P** coincides with the analytic route between P* and P**. So P and P* are identical exactly if they have the same analytic profile.

How are we to reconcile this analysis with the, from some viewpoint, very reasonable idea that a proposition is the set of its consequences? How are we to understand the concept of an "analytic profile" in terms of the idea that a proposition is the class of its consequences?

A proposition, I would suggest, is a *structured* set of its consequences. *Triangles have three sides* and *1+1=2* are sets that have *ordinal structure.* Those sets comprise exactly the same propositions. But the way in which they are ordered in the one set differs from the way in which they are ordered in the other. That is why those propositions have different analytic profiles and, therewith, entirely different cognitive values. I work this out in detail in my book (unpublished) *Conceptual Atomism and the Computational Theory of Mind.*

A problem with our analysis is that it seems viciously circular. Surely P is one of its own logical consequences. So is *P or Q* and *P or (Q or*

R), and so forth. Now if we analyze P as a conjunction, among whose conjuncts it numbers, then surely our analysis of P is viciously circular. (Some versions of non-standard set theory permit such circularity. But surely we don't want the viability of our analyses of propositionality to depend on the viability of such controversial doctrines.)

This problem can be finessed. Obviously some restriction is needed on *which* consequences of P we are to include in its content. The right restriction, I believe, is something like this. P is the class of its "proper" consequences, where q is a *proper* consequence of p iff q is not some kind of molecularization of p, i.e. if q is not built up out of p. So *x is colored* is a proper consequence of *x is blue*, whereas *x is green or blue* is not. The details, and also the surrounding theory, are worked out in *Conceptual Atomism and the Computational Theory of Mind.*

[113] Kaplan 1989.

[114] I borrowed this expression from Blackburn 1984.

[115] Except, of course, where analytic thoughts are concerned.

[116] I am alluding to the controversial question whether all information is embodied, or is capable of being embodied, in propositions. Sense-perceptions are analogue; propositions are digital. So, on the face of it, it seems that the kind of information mediated by sense-perceptions would be non-prepositional. MacDowell (1993) famously argues against this. I also argue against it. ("Some Problems for Intentionalism". *Acta Analytica.*) We don't need to settle this issue here. Here all that is relevant is that hallucinations tell you something false and, consequently, hallucinations are proposition-like *in that* nothing needs to be added to them to make them either true or false: they are, apparently, "saturated". So a hallucination is not something with a blank in it. A veridical perception is not something that is like a hallucination except that the blanks have been filled in. Given a veridical perception P, and a hallucination H that is qualitatively just like P, both P and H are equally "saturated". So P is not what results when something has been added to the *content* of H.

[117] I am using the term "information" to denote what the perception *tells* me. Many writers, most notably Dretske, say that the "information" borne by a perception, or anything else (e.g. a photographic plate), consists in the states of affairs that stand in a certain causal relation to that perception. I think this is misleading. The "information" borne by a perception is, truistically, what the perception tells you. Obviously there is a relation

between "information" in my sense and "information" in Dretske's sense. But I think it only leads to confusion to use the term "information" in the latter sense. The word "information" has mentalistic overtones. If, by stipulation, we say that the "information" borne by x consists in those states of affairs that bear a suitable causal relation to x, then we end up prejudging, through use of terminology, important issues regarding mental representation and information (in the sense of what your perceptions *tell* you, not in Dretske's sense).

What you can *deduce* from a perception p -- in light of your background knowledge of causal laws, uniformities, concomitances and the like – is different from what the perception *by itself* tells you. The former consists of (inter alia) those states of affairs which stand in a suitable relation to p. The latter is not as expansive. This is a point made (in a different way, in a different context) by Sellars in *Empiricism and the Philosophy of Mind.* Jones is in a tie-shop where the highly unusual lighting conditions completely scuttle the correspondence that usually holds between one's visual sensations and the colors that objects have. By itself, Jones' perception of the tie doesn't tell him that it is red. But *given* a backlog of environment knowledge, Jones can deduce, on the basis of his perception, that he is seeing a red tie. What I am saying is: when we focus on what a perception *by itself* tell us – when we don't telescope into it a wealth of background knowledge – we remove one of the supporting pillars of content-externalism.

[118] See McGinn 1988.

[119] See Blackburn 1984, Loar 1988.

[120] See Loar 1988. For powerful criticism's of this kind of view, Burge 1982.

[121] This move seems to be in the spirit of Burge (1982).

[122] Putnam 1975.

[123] See Burge 1982.

[124] See Kripke 1972.

[125] See Burge 1982.

[126] Kripke himself confines himself to saying only that they are not descriptions; he does not explicitly say that they are mere labels. I myself don't see any third option. some have argued that names have both a "primary" and a "secondary" meaning, and that Kripke is right about one of those meanings, but wrong about the other – so an attenuated descriptivism holds. Scott Soames (2005) admirably criticizes these views. My own view is

that we should leave semantics alone, as much as possible: when we encounter prima facie reasons to be descriptivists regarding names, we should try to explain away those counter-examples by reference to epistemology – to facts relating to our *grasp* of literal meaning. This is the course I've been pursuing in the present work. Given what Kripke said, it is very hard to make a case that, at any semantic level, "Socrates" and "Hesperus" are descriptions: and, as we have seen, it isn't necessary to make that case to deal with the data that motivate descriptivism.

[127] Kripke 1972: 141-142. Quoted in Soames 2005: 73-74.

[128] See Stroud 2001.

[129] I believe that Wittgenstein denies it in his *Remarks on the Foundations of Psychology*. With all due respect to that great author, his grounds for this *prima facie* highly counter-intuitive position are, at least to me, very obscure.

[130] Jerry Fodor 1998 has gone so far as to say that a causal connection to x is *all* that is involved in having a concept of x. So I can have *no* beliefs about the number and yet have a conception of it: I need only have a casual connection to it (or, more accurately, some instance of it). I need only have a causal connection to Socrates to be able to think about him. I needn't know *anything* about him.

[131] If I'm not mistaken, what Kripke says is not so much that people know of *no* individuating descriptions of the things they think about and refer to. His point is that the individuating descriptions are not the ones that Russell (and possibly Frege) *thought* they were. I think that Kripke himself says something similar to what we say below – when we argue that one *does* need (knowledge of) an individuating description *of a very special kind* to refer to an object. (I won't press these exegetical points. I will argue only for the point that none of Kripke's cogent points to any degree warrant the rejection of the "descriptivism" I have been urging.)

[132] What I will say goes a bit beyond what Kripke says; my own commentary is interlaced with an exposition of Kripke's views.

[133] More exactly, that person gives Socrates the Ancient Greek equivalent of "Socrates", whatever that is. In this discussion, we will ignore the fact that, over the years, names may undergo considerable phonemic change. Obviously "Confucius" sounds very little like the name of the old

sage. These phonemic changes introduce no philosophical mysteries, but allowing for them would clutter up the exposition considerably.

[134] Of course, the term "occasion" has to be defined somewhat narrowly. One can refer to twenty different people in a single conversation and thus on a single "occasion". But at any *specific* juncture of that occasion – at any *one* time (leaving aside ambiguity) – one is referring to just one person or object. When I say "occasion" I mean a specific juncture of a discourse.

[135] See Russell 1948 (89-99).

[136] It isn't even clear if Russell *was* wrong about the nature of the uniquely individuating descriptions. It could be that he was simply *using* historically pregnant descriptions to illustrate his basic point, which is that one must have *some* uniquely individuating description of Socrates to think about, or refer to, Socrates. If he had illustrated his point with some uniquely individuating description that didn't *obviously* apply to Socrates – e.g. *the guy to whom Seymour was referring on occasion C…*-- nobody would have followed what he was saying. So his use of uniquely individuating descriptions like *the philosopher who drank Hemlock* may not really indicate his belief that *that* is an actual uniquely individuating description through which one might refer to, or think about Socrates. It might just be a superficial expository device. But I am not a Russell scholar; so I remain open on this.

[137] See Kaplan 1968, 1989a, 1989b, and Putnam 1975.

[138] See Jackson and Pettit 2004b-2004e, especially 2004b (58-62), 2004d (105-109), and the whole of 2004e.

[139] Perhaps one might distinguish *mental* content from *representational* content. One might say that Smith and twin-Smith have qualitatively identical mental contents, but different representational contents. There are two things to say. First of all, it seems at best arbitrary to separate so sharply between the mental and the representational. But suppose *arguendo* that we grant that distinction. In that case, by the argument just given, externalism strips representational content of causal power – not a palatable result.

[140] See Jackson and Pettit 2004b-2004e, especially 2004d (108-114) and 2004b (58-62).

[141] The same is true of most of their applications of that concept.

Particularly impressive are their applications of it to social theory (see Jackson and Pettit 2004f) and to the concept of causal explanation itself (see Jackson and Pettit 2004g). For example, they argue (2004f) that in *some cases*, it is more informative to know a *program* cause than an actual cause, since the former is richer in counter-factual information. Thus, in some cases, a macroscopic course-grained explanation is actually *better* than a microscopic, fine-grained explanation. To my astonishment, I found myself in complete agreement with this radical point.

[142] See Jackson and Pettit 2004b-2004e, especially 2004c (48-49).

[143] Thanks to Dan Robinson for suggesting that I write and include this chapter.

[144] Dummett 1973.

[145] See Soames 2001 for an excellent discussion of Dummett's position.

[146] Kripke 1974.

[147] This is what Kaplan 1989 says.

[148] I owe this point to Tim Stowell, department of linguistics University of California, Los Angeles, and to Susanne Cummins, department of linguistics University of California, Santa Barbara.

[149] Of course, some will deny this:

> It is patently obvious that the concept *horse* (or *x is a horse*) is semantically contributed by the term "horse" in (H).

I must disagree. Suppose that Alpha is the only contextually salient object. I point to it and say:

(H2) "that is a lovely beast"

We know that (H2) encodes the proposition

(AL) *Alpha is a lovely beast.*

Given this, suppose we say, along with the objector, that (H1) encodes some proposition comprising the concept *x is a horse*. What would that proposition be? How can we make the concept *x is a horse* be a

constituent of the literal meaning of (H1)? To do this, we must suppose that (H1) encodes some kind of quantified generalization. Let D be the domain of discourse delimited by the ostension. In that case, the semantic content of (H1) becomes either:

for any x within D, if x is a horse, then x is a lovely beast

or

for some horse x within D, x is a lovely beast.

Remember what we said earlier in connection with "rhenates" and "chordates". Whenever a noun-phrase contains a sortal, if you make that sortal be part of the semantic contribution of that noun-phrase, the result is that the sentence in question encodes a generalization of some kind. So if the objector is right, and the concept "horse" is semantically contributed by the occurrence of "that horse" in (H1), then (H1) and (H2) have utterly different semantics. One would have a singular proposition for its meaning; the other would encode a universal quantification. This is simply not feasible. There is no doubt that *both* (H1) and (H2) encode singular propositions: in each case, the literal meaning consists exactly of the attribution, to some object, of the property of being a lovely beast. In both cases, obviously, the object is Alpha. So (H1) and (H2) both encode (AL): nothing more, nothing less.

[150] This is a version of Church's famous "translation" argument.

[151] Clarence Lewis (1946) makes this point.

[152] There is one very subtle point we must attend to. Again consider:

(^) "John owns a donkey. He beats it often."

I said that the "he" refers to the thing described by the previous sentence. A problem is that the first sentence might not give a *uniquely individuating* description. (^) could be true if John had several donkeys. (Suppose John has ten sons, one of whom he loves. I can still correctly say: "John has a son. He loves him.")

We do, of course, refer to things that are merely described. But the descriptions have to be uniquely individuating. If John has ten donkeys, then what ensures that the "he" will refer to the right one? So, it might be

argued, since (^) doesn't give us a uniquely individuating description, the "he" in it is not a singular term; it doesn't refer. This would suggest that Neale's analysis is preferable to ours.

I would suggest a different approach. I agree that, if we are to refer to a thing merely described, we need a uniquely individuating description. But I think that, in (^), we ultimately have one. Suppose your swashbuckling friend Bob tells you the following tale of woe:

"I was once in love with a woman. She broke my heart. Although I've been in love ten times since then, it's never really been the same…"

The obvious (but false) thing to say here is: The description in the first sentence is indefinite. Bob has been in love with many women. Therefore the existence claim in question doesn't single anyone out: there is no one who *uniquely* satisfies it.

But when Bob produced that initial sentence, he had some *particular* woman in mind. So *in effect* the existence claim in question is: *there is some woman x whom Bob has in mind on occasion O and such that Bob was in love with x…* That existence claim *is* uniquely individuative: there *is* one only woman whom Bob has in mind on that particular occasion.

Suppose now that Bob utters (^) to you. It is true that that many donkeys may satisfy the description *donkey that John owns.* But there is some *one* donkey x such that the person stating (^) has x in mind. So the relevant description (*thing x such that the person uttering (^) has x in mind*) is uniquely individuative.

[153] See Blackburn 1984, Neale 1990.

[154] See Grice, "Logic and Conversation". Reprinted in *Philosophy of Language*, edited by A.P. Martinich. Oxford University Press: 1990.

[155] Donnellan 1966.

[156] Actually, A.J. Ayer made exactly that distinction around ten years before Donnellan. It occurs in the last chapter of his book *The Problem of Knowledge.*

[157] See Kuczynski 2005.

My Spanish is a bit rusty. This needs proofing.